Universidad Autónoma de Madrid

Departamento de Biología Molecular

Facultad de Ciencias

STRUCTURAL STUDIES ON *E.coli* DNA POLYMERASE III CLAMP LOADER SUBCOMPLEXES

PhD by
Marta Rajkiewicz

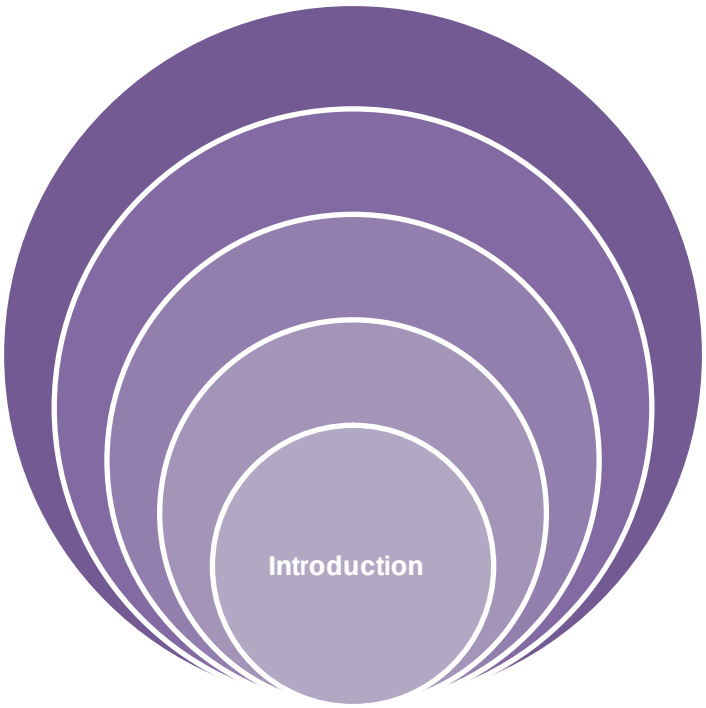
Under direction of:
Dr. Carmen San Martín Pastrana
Prof. José María Carazo García

Abstract

A crucial step in DNA replication is opening of the ring-shaped sliding clamp, and its proper placement on the DNA template. This process is carried out by the clamp loader complexes, in an ATP dependent reaction that is stimulated by the primed DNA. In the Gram-negative bacteria *E. coli*, two main components of the clamp loader are γ and τ, products of the same *dnaX* gene, that isolated in solution, form homotetramers. In the presence of additional subunits δ and δ', the equilibrium shifts and the $(DnaX)_3\delta\delta'$ is formed. Other two subunits, ψ and χ, form a complex with the clamp loader. The ψ subunit binds to the DnaX proteins (τ or γ) on one side and to χ on the other. The ψ subunit forms a bridge between $(DnaX)_3\delta\delta'$and χ links the clam loader to SSB facilitating the displacement of primase from newly synthesized RNA primer, thus playing a crucial role in Okazaki fragment synthesis.

We have analyzed the structure of *E. coli* γ_4 and $\gamma_4\psi\chi$ complexes using three-dimensional electron microscopy. Reaching convergence in 3D map refinement was hindered by several reasons: the small size (165 kDa to 222 kDa), lack of symmetry, and globular shape of the complexes, which resulted in low signal in the images; and large structural heterogeneity, which could be caused by intrinsic complex flexibility, and was refractory even to sample cross-linking procedures. Nevertheless, by using maximum-likelihood image classification methods we were able to select homogenous subpopulations of particles and calculate consensus three-dimensional maps for all complexes. Combination of our maps with crystallographic or homology model structures show how DnaX proteins can establish a large variety of interactions among them, in agreement with their versatility in forming complexes with different accessory proteins during polymerase holoenzyme assembly. Difference maps show the location of the ψχ heterodimer in the $\gamma_4\psi\chi$ complex, and how the interactions among γ subunits are altered by its presence.

Contents:


Introduction

1. *Introduction*

1.1 DNA replication

DNA replication is one of the most important processes in the living organism. All organisms need to duplicate their genetic material that contains the genetic code. Replication is carried out in a semiconservative manner: from one mother molecule of DNA, two daughter ones are replicated, each containing one original and one new strand. The leading strand is synthesized going along with the fork movement, while the lagging strand runs in the opposite direction and is synthesized in a piece fashion (as Okazaki fragments). Various proteins cooperate to assemble the two new DNA chains in a highly coordinated way, approaching speeds of 1000 nucleotides per second (in *E.coli*). A useful analogy to illustrate the complexity of DNA replication was proposed in (Baker and Bell, 1998). If we assume that the DNA duplex has a diameter of 1 meter, the replisome machinery would be a formula-1 car circulating at 600 km/h speed. To replicate the whole genome, the circuit length would be about 400 km and the race would take about 40 min, with the possibility of our car getting lost (or getting out of the race) each 170 km.

The process of DNA replication is divided into three general steps: initiation, elongation and termination. The fundamental players involved in DNA replication are conserved from viruses to bacteria, archaea, and eukaryotes. Since *E. coli* is the experimental system where DNA replication has been most extensively studied, we will use the *E.coli* replication protein names to describe the replication process in general terms (Kornberg, 1992). In the initiation stage, protein DnaA binds to the *E. coli* replication origin, oriC, melting an AT rich region to form a single-stranded DNA bubble. When the replication bubble is created, the replisome is assembled and the replication fork is formed, to proceed to the elongation stage. Key components of the *E.coli* replisome (Figure 1) are:

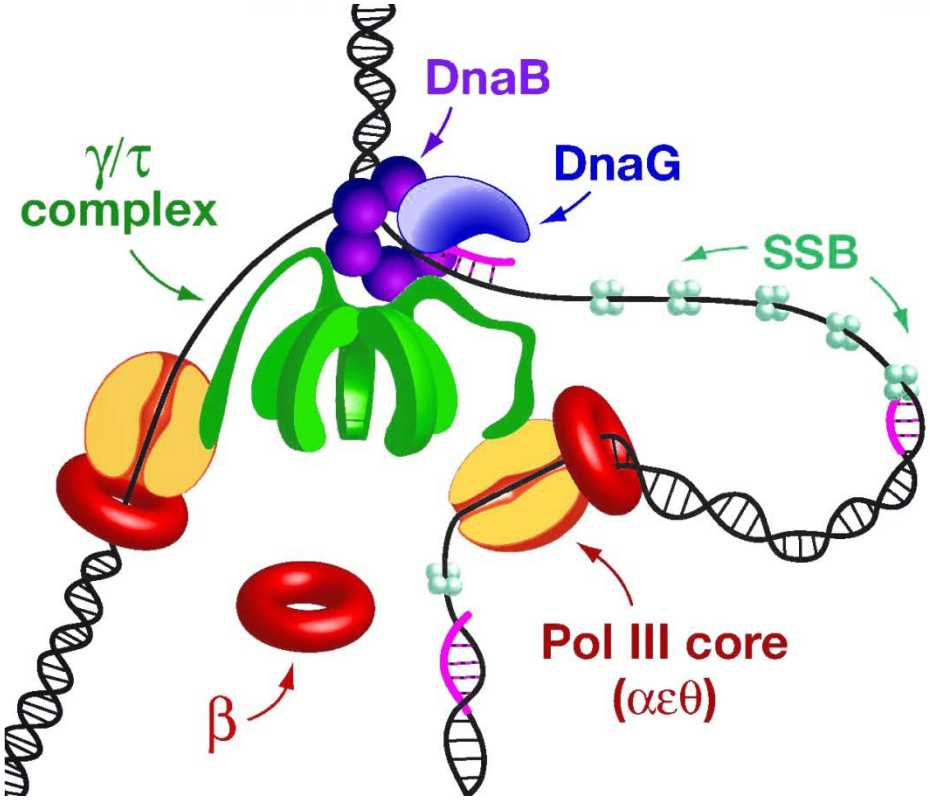


Figure 1. Organization of the *E.coli* replisome (Johnson and O'Donnell, 2005)

- Helicase: a hexamer of DnaB protein that separates the duplex DNA into two single strands by translocation in the 5′ to 3′ direction along the lagging strand, using the energy of ATP hydrolysis. DnaB is placed in the replisome by its loading partner, DnaC.
- SSB: Single-stranded DNA-binding proteins coat each individual strand to remove DNA secondary structures that would impede replication.
- Primase: DnaG primase catalyzes the synthesis of short RNA fragments, called primers.
- DNA polymerase: Elongation of RNA primers proceeds simultaneously on both strands by the action of the replicase, DNA polymerase III holoenzyme (Pol III). Pol III is the only polymerase present in the *E. coli* replisome. The polymerase uses the single-stranded DNA as a template to synthesize a complementary strand. However, this is not the only function performed by Pol III, as will be described in the next section.

E. coli chromosomal DNA replication proceeds bidirectionally from the origin, and the two replication forks meet in the terminus region, which contains ten 23-bp

Ter sites, to which the replication terminator protein binds (Goedken, Levitus, Johnson, Bustamante, O'Donnell and Kuriyan, 2004). The fork that arrives first in the terminus region is halted by its encounter with a correctly-oriented complex of the Tus replication terminator protein bound at one of several *Ter* sites, where it is trapped until the other fork arrives. This mechanism is called a 'replication fork trap' (Duggin, Wake, Bell and Hill, 2008).

1.2 E. coli replicase: DNA polymerase III

Replicative DNA polymerases, or replicases, play a key role during elongation. In *E. coli* the replicase role is performed by the DNA polymerase III holoenzyme (Pol III). The Pol III holoenzyme is formed by 10 different polypeptides, carrying out different functions (Table 1). The catalytic core, composed by subunitsα, ε and θ, contains the polymerase and exonuclease enzymatic functions. Subunit α is in charge of adding new nucleotides to the nascent DNA strand. If Pol III mistakenly incorporates an incorrect nucleotide (error rate of approximately 10^{-5} to 10^{-6} per base replicated), the proofreading ′3to 5′ exonuclease subunit ε removes the incorrect nucleotide. One catalytic core is bound to each replicating strand (Figure 1).

Ring-shaped subunit β acts as a clamp, tethering the catalytic cores to the replicating strands (one ring per catalytic core) and ensuring processivity. In the lagging strand, Pol III translocates along the DNA in the opposite direction of the Pol III complex on the leading strand and the whole replisome at the replication fork (i.e., fork progression). Moreover, the catalytic core in the lagging strand has to “hop” on and off every time a new Okazaki fragment is initiated or finished. All this complicated process is coordinated by a multiprotein complex known as the clamp loader. ATPase subunits in the clamp loader (τ or its shorter version, γ) use the energy of ATP hydrolysis to open the β clamp and subsequently to close it around the DNA. The τ subunit contains flexible C-terminal extensions that organize the replisome by simultaneously binding the Pol III catalytic cores and the DnaB helicase (Figure 1).

Because two polymerases are required at the replication fork (one for the leading and another for the lagging strand) it has long been assumed that the Pol III active form was a dimer of calaytic cores, coordinated by one clamp loader. However, new

evidence indicates that three catalytic cores are present at the *E. coli* replisome, both *in vitro* and *in vivo* (Blinkova, Hervas, Stukenberg, Onrust, O'Donnell and Walker, 1993; Lovett, 2007; McInerney, Johnson, Katz and O'Donnell, 2007; Reyes-Lamothe, Sherratt and Leake, 2010).

Table 1. Subunits of *E.coli* DNA polymerase III holoenzyme and their functions (courtesy of Charles McHenry).

Subassembly/Subunit			**Gene**	**Length (aa)**	**M.W. (kDa)**	**Function**
Pol III*	*Polymerase*					
	catalytic core	α, alpha	*dna*E	1160	129, 905	Polymerase
		ε, epsilon	*dna*Q	243	26,907	3'→5' proofreading exonuclease
		θ, theta	*hol*E	76	8,788	binds and stabilizes ε—not essential
	Clamp loader					
	DnaX proteins	γ, gamma	*dna*X	430	47,414	ATPase, contacts δ, δ' and ψ through domain III
		τ, tau	*dna*X	643	71,007	functions listed for gamma plus interaction with α (through domain V); multimerizing polymerase and interaction with DnaB helicase (through domain IV)
	δ', delta'		*hol*B	334	36,937	required for loading β, binds DnaX
	δ, delta		*hol*A	343	38,704	contacts β and required for β loading, binds δ'
	χ, chi		*hol*C	147	16,599	binds SSB, confers salt resistance on elongation
	ψ, psi		*hol*D	137	15,043	Binds DnaX and together with χ increases affinity of DnaX for δ-δ' [check is this is still true after MS paper]
	Processivity Factor					
	β, beta		*dna*N	366	40,587	processivity factor, encircles DNA and tethers polymerase by protein-protein interactions

1.3 The E. coli clamp loader

Clamp loaders are generally heteropentameric complexes with sequence similarities among their subunits. They are needed to coordinate, communicate and organize the various replication proteins. The *E. coli* clamp loader contains 6 different subunits: two active ATPases, γ and τ, encoded by the same gene *(dnaX*); δ and δ'; ψ and χ.

τ is the full length translation product of the *dnaX* gene, and γ is a smaller version truncated by a translational frameshift (Blinkowa and Walker, 1990; Flower and McHenry, 1990; Tsuchihashi and Kornberg, 1990). Both τ and γ contain an ATPase domain (domains I and II) and a domain responsible for oligomerization and interaction with the δ, δ' and ψ clamp loader accessory subunits (domain III). The difference in size between γ (47 kDa) and τ (71 kDa), is a 24 kDa C-terminal region (domains IV and V) that allows τ to bind to the DnaB helicase and to the α subunit in the Pol III core (Figure 2) (Dallmann, Kim, Pritchard, Marians and McHenry, 2000; Gao and McHenry, 2001b; Kim, Dallmann, McHenry and Marians, 1996). As a result of these interactions, τ multimerizes the Pol III core and coordinates the leading and lagging strand synthesis (Gao and McHenry, 2001a; Gao and McHenry, 2001b; Gao and McHenry, 2001c).

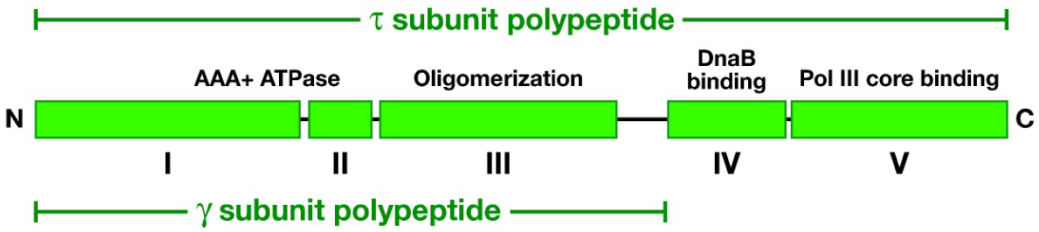


Figure 2. Domain organization of *E. coli* γ and τ (Johnson and O'Donnell, 2005).

Although δ and δ´ present sequence and structure motifs characteristics of the AAA+ ATPase family, they do not bind ATP. They perform different functions in the clamp loader. δ acts as a wrench, opening the sliding clamp to position it onto the replicating DNA strand (using energy from protein-protein interactions) . δ´ is a stator protein, In contrast to *γ* and *δ*, the domains of have more intermolecular interactions possessing

the same orientation as in δ , which led to name subunit, “stator,” being the stationary part of a machine upon which other parts move. The conformation adopted by γ1 appears to be prevented in δ′lack ing a flexible linker between Domain II and Domain III cannot perform the conformation that does γ1 .What is more, the presence of hydrophobic interactions at the II-III interdomain interface ′inlimδit conformational changes at the interface. The relatively closed “C” shape of δ in isolation (Guenther, Onrust, Sali, O'Donnell and Kuriyan, 1997) is maintained in the crystal structure of the γ complex (Johnson and O'Donnell, 2005). When ATP binds to the DnaX gene product, the clamp loader undergoes a conformational change exposing the β-binding site in δ; binding occurs and the ring opens. When ATP is hydrolyzed, conformational changes in the DnaX subunits move δ towards δ`, for closing the β ring (Davey, Jeruzalmi, Kuriyan and O'Donnell, 2002). ATP hydrolysis and ADP release destabilize the δ-β interaction, thus releasing the clamp to continue the replication cycle.

Both χ and ψ are monomeric in solution, and form a 1:1 hetrodimer. ψ forms insoluble aggregates, and so it has to either be purified in denaturing conditions and solubilized in the presence of χ (Xiao, Dong and O'Donnell, 1993), or jointly expressed with χ in a single operon (Olson, Dallmann and McHenry, 1995). Neither ψ nor χ are needed for loading of the clamp, but they play a crucial role in Okazaki fragment synthesis. ψ forms a bridge between the $DnaX_3$δδ` complex (via interaction with the domain III of DnaX) and χ provides conformational stability to the complex (Gao and McHenry, 2001c). χ interacts with the SSB proteins when bound to DNA. The interaction between χ and SSB facilitates the displacement of primase from the newly synthesized RNA primer, thus being freed to be recycled to another side (Glover and McHenry, 1998; Gulbis, Kazmirski, Finkelstein, Kelman, O'Donnell and Kuriyan, 2004).

In solution, both *dnaX* gene products form tetramers. This was first determined by sedimentation equilibrium analytical ultracentrifugation assays (Dallmann and McHenry, 1995). More recently, mass spectrometry (Anderson, Williams, O'Donnell and Bloom) assays have confirmed and added detail to these results (Park, Jergic, Politis, Ruotolo, Hirshberg, Jessop, Beck, Barsky, O'Donnell, Dixon and Robinson,

2010). MS showed that γ_4 is more stable than τ_4: while both are predominantly tetrameric in solution, smaller oligomers were found for τ, but not for γ, in the same ionic strength conditions (Figure 3) . The equilibrium shifts to a trimer when δ and δ´ are present, forming a $DnaX_3\delta\delta`$ complex (Pritchard, Dallmann, Glover and McHenry, 2000). δ´ acts as an oligomer breaker, directing the transition from from the τ/γ tetramer to the trimer (Park, et al., 2010). Both DnaX forms can produce well defined complexes with δ, δ´ and ψχ, ($\tau_1\gamma_2\delta\delta`\psi\chi$ or ($\tau_2\gamma_1\delta\delta´\psi\chi$) (Pritchard and McHenry, 2001). The ψχ heterodimer can form complexes with the DnaX homo-oligomers in the absence of δ or δ´ (C. McHenry, personal communication, and (Park, et al., 2010)). MS assays indicate that although ψχ has a limited oligomer breaker activity, it preferentially associates with the tetrameric, rather than the trimeric, DnaX form, to produce $DnaX_4\psi\chi$ complexes (Figure 3c) (Park, et al., 2010).

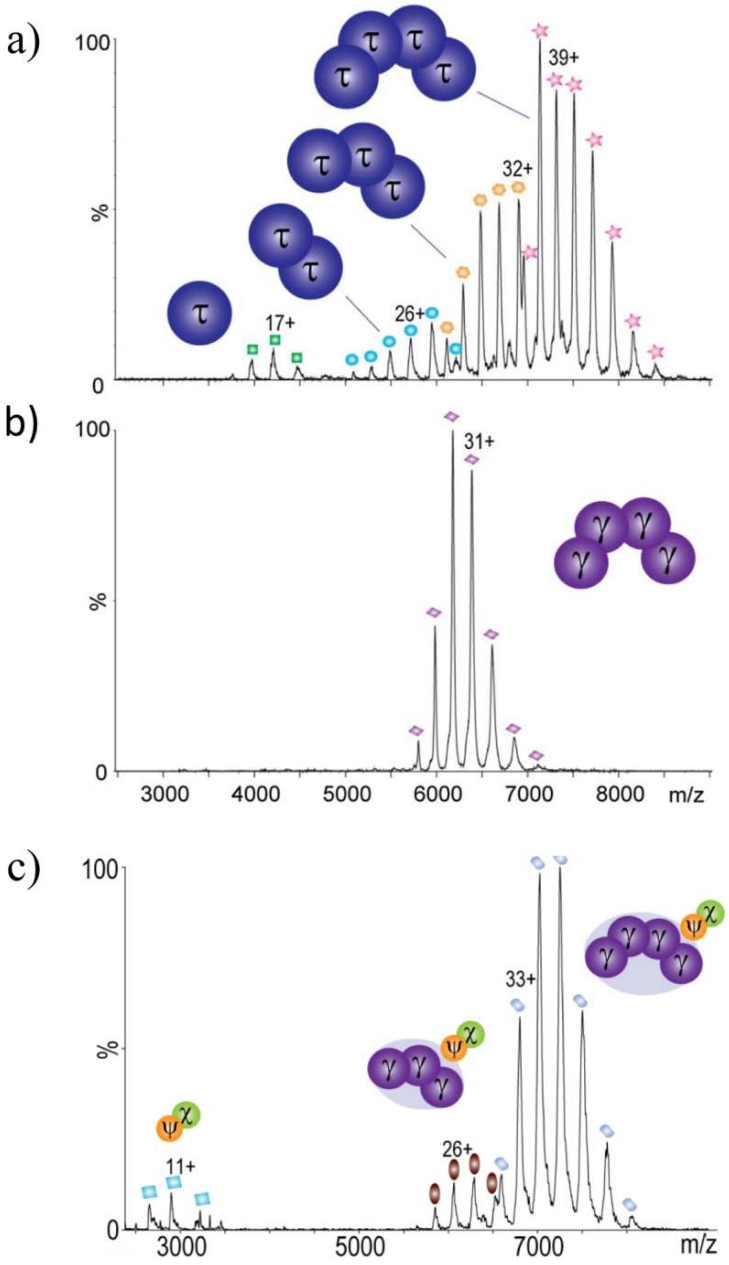


Figure 3. Mass spectra of: a) τ and b) γ, both in 0.1M NH4OAc at pH 6.9 low ionic strength buffer; and c) γ when ψχ is present (Park, et al., 2010).

1.4 Structural knowledge on the E. coli clamp loader

The $\gamma_3\delta\delta'$ complex is considered the minimal clamp loader, as only those subunits are absolutely required for β clamp placement. The crystal structure of the *E. coli* minimal clamp loader has been solved in the absence of any cofactor (Jeruzalmi, O'Donnell and Kuriyan, 2001), as well as in the presence of ADP and the non-hydrolizable analog ATPγS (Kazmirski, Podobnik, Weitze, O'Donnell and Kuriyan, 2004). All structures are identical, and contain three copies of γ truncated to residue

373 (the whole sequence is 430 residues) as well as δ and δ´, assembled in a heteromeric pentamer of 200 kDa molecular weight. A truncated version of γ was used for the crystallization, since the remaining C-terminal residues (374-430) are highly sensitive to proteases (Jeruzalmi, et al., 2001) and predicted to be unfolded (Prilusky, Felder, Zeev-Ben-Mordehai, Rydberg, Man, Beckmann, Silman and Sussman, 2005). Each of the subunits has three domains: an N-terminal domain with a RecA-like fold (domain I), and two helical domains (domains II and III). They are arranged in a two-tiered asymmetric ring. One of the ring tiers is formed by the C-terminal domains (III) forming a closed collar that holds the pentamer together, as there the strongest intersubunit contacts occur. The N-terminal domains (I, II) form the other tier, which is open with a gap between the δ and δ´ subunits (Figure 4). Modeling studies showed that domain I of δ must be rotated outwards by 20-30^{o} with respect to domain III, for proper interaction with β (Leu and O'Donnell, 2001).

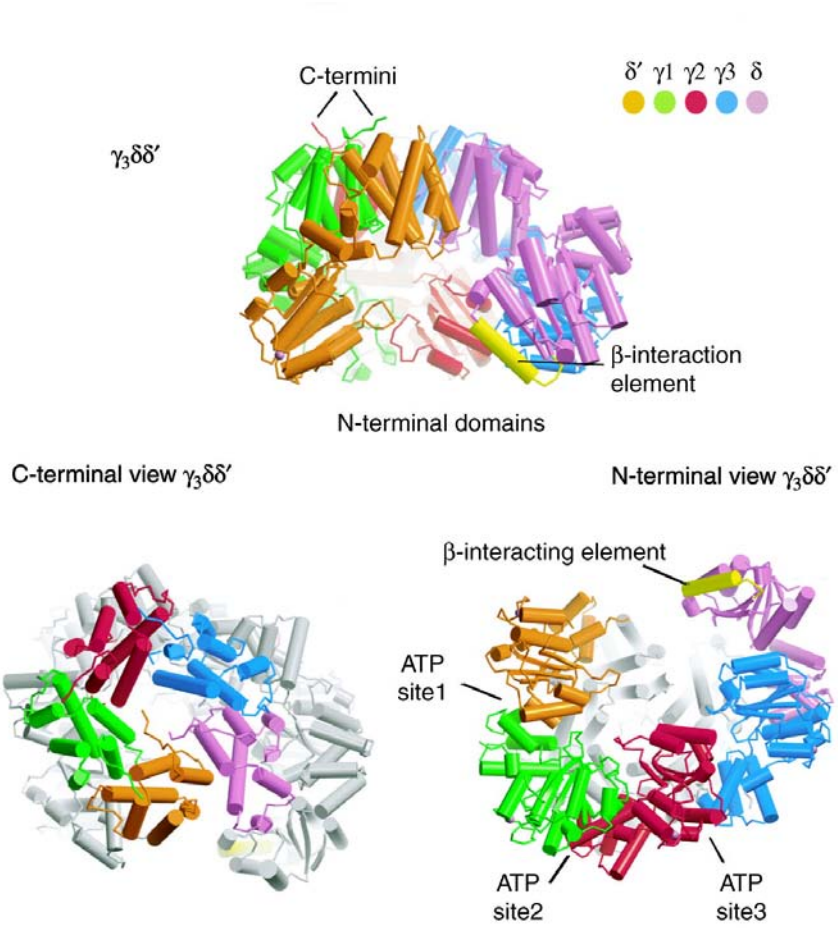


Figure 4. Minimum clamp loader structure $\gamma_3\delta\delta'$; its C and N-terminals views (O'Donnell, Jeruzalmi and Kuriyan, 2001).

The most important information that we can take from this crystal structure concerning the present work is that each γ subunit adopts a different conformation, although they are identical polypeptides. When viewed from above the C-terminal collar, γ3 domain I is rotated outward the most (Figure 4). If we superimpose domains II of all three γ subunits in the clamp loader, the relative rotations between their domains III are: 21º for γ1-γ2; 45º for γ2-γ3; and 60º for γ3-γ1 respectively (Figure 5) (Jeruzalmi, et al., 2001).

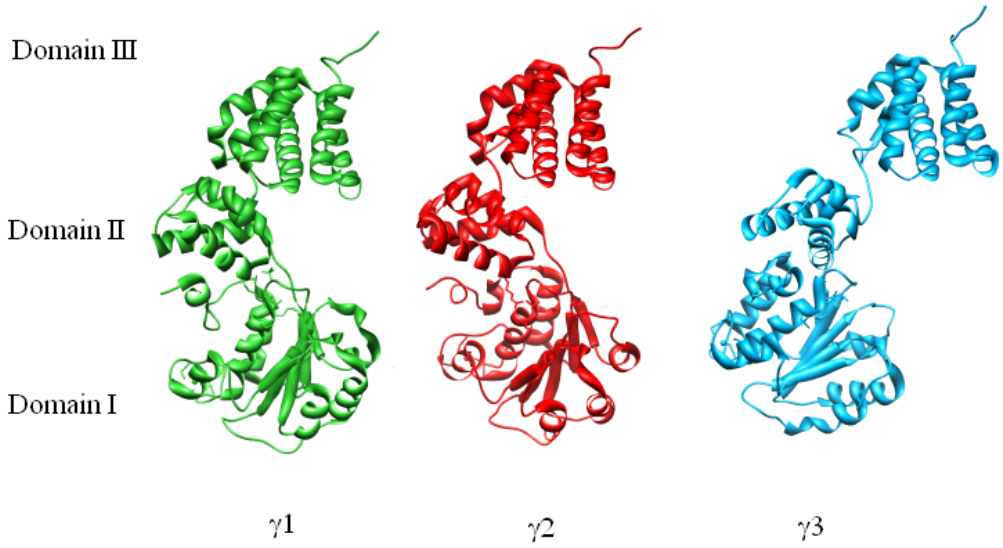


Figure 5. Individual γ subunits from the $\gamma_3\delta\delta'$ crystal structure. The three domains (I (N-term), II, and III (C-term)) of each subunit are indicated.

The crystal structure of the ψχ heterodimer has also been solved (Gulbis, et al., 2004). The final crystallographic model lacks the first 26 residues at the N-terminus of ψ (Figure 6a). The N-terminal region of ψ, responsible for binding to the C-terminal domain of γ, was disordered and could not be traced. It was proposed that exploiting the flexibility of the ψ N-terminal region, the ψχ dimer could be placed in the gap between δ and δ´ in the clamp loader(Figure 6b).

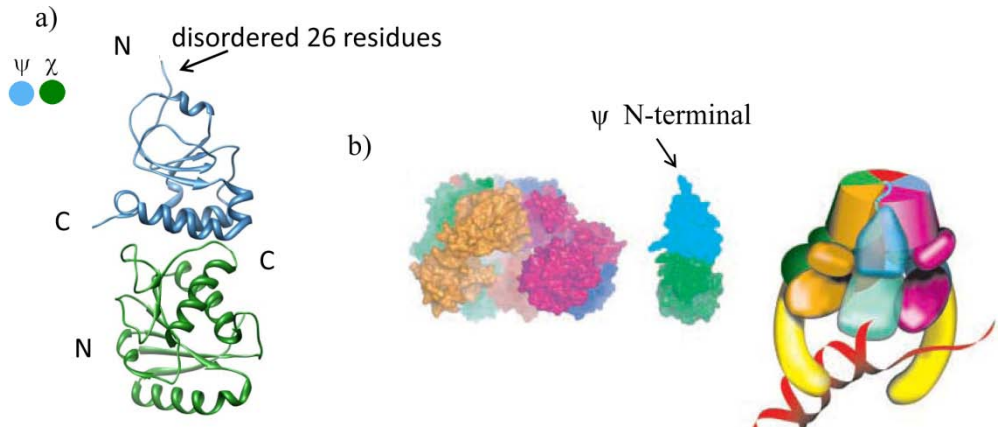


Figure 6. (a) Crystal structure of the ψχ heterodimer (b) Scheme of the $\gamma_3\delta\delta'$ clamp loader and possible filling of the space between δ and δ′ by ψχ. Clamp loader subunits are colored as in Figure 4 and 5 (Gulbis, et al., 2004).

More recently, the crystal of the minimum clamp loader bound to primer/template DNA and a peptide comprising residues 2-28 at the N-terminus of ψ has been solved (Simonetta, Kazmirski, Goedken, Cantor, Kelch, McNally, Seyedin, Makino, O'Donnell and Kuriyan, 2009). The ψ peptide binds on top of the C-terminal collar region, interacting with domains III of all three γ subunits (Figure 7). Based on this structure, the model in which the ψχ dimer fills the clamp loader gap does not seem to be correct, as the C terminus if the ψ peptide points away from it. The presence of the ψ peptide distorts the C-terminal collar formed by domains III. The C-terminal domain of the γ_3 subunit (the closest to the δ wrench) rotates towards its domain II to make space for entry of the ψ subunit. That is, a fourth conformational state of γ is revealed, emphasizing the molecule flexibility and its possible relation to the different functional states in the replicase.

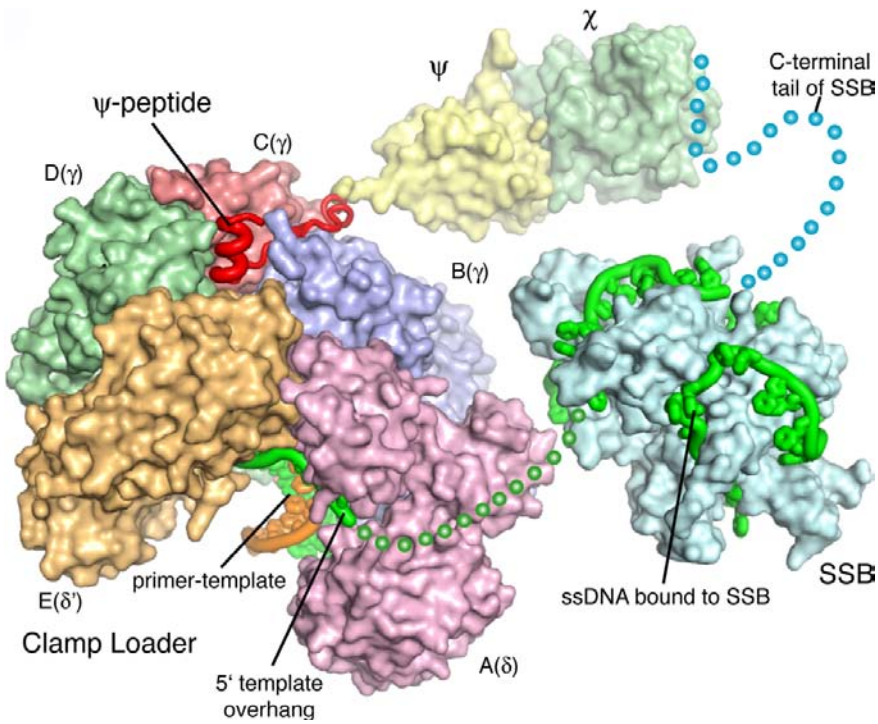


Figure 7. Model for the possible arrangement of the ψχ dimer in the active replisome. It is based on known structures of the ψχ heterodimer and the γ3δδ` clamp loader (Simonetta, et al., 2009).

1.5 3DEM: the sample

Since 20 amino acids can be combined to form a protein in many different ways, the three-dimensional structure cannot be predicted. It has to be determined experimentally using techniques such as crystallography, NMR (Nuclear Magnetic Resonance) or 3DEM (Three Dimensional Electron Microscopy). The latter is the technique used in this thesis. Each of these techniques has its own requirements and provides different and complementary information. The most important difference between them, is the conditions of the sample to be used (Table 2). NMR needs proteins in solution; crystallography needs an extremely pure and very concentrated specimen, forming well ordered 3D crystals. 3DEM can work with smaller amounts of protein, and small impurities are allowed as far as they are distinguishable from the structure of interest in the micrograph. 3DEM is the most versatile of the three regarding sample size, as it can cover a quite ample range between 10^2 and 10^9 Da. Additionally, 3DEM can work with flexible complexes up to a certain degree (Chiu, 1993; Frank, 2006). Analysis of macromolecular flexibility and how to deal with heterogeneity was always a challenge. There are some methods and basic principles that help to solve this problem, such as: combination of X-ray crystallography,

electron crystallography or NMR to obtain pseudo–atomic models, 3D multivariate statistical analysis generally used for cro-EM due to requirement of large data set, supervised 2D and 3D classification (bootstrap method), 2D and 3D maximum likelihood methods both in real and Fourier spaces, adding of tilt angles (Random Conical Tilt) or multi-reference refinement (Leschziner and Nogales, 2007; Spahn and Penczek, 2009).

Table 2. Difference between sample quantities and concentrations in cryo-EM, crystallography and NMR techniques (Frank, 2006).

Techniques	**Cryo-EM**	**X-Ray**	**NMR**
Sample volume	5 µl	10 µl	400 µl
No. of samples/experiment	5–10	1 crystal	1 (<20 kDa), several (20–40 kDa)
Concentration	50 nM	50 µM	0.5–1.0 nM

3DEM uses as basic equipment a transmission electron microscope (Chen, Settembre, Aoki, Zhang, Bellamy, Dormitzer, Harrison and Grigorieff). The image is formed by focusing an electron beam with electromagnetic lenses. The electrons would be scattered by molecules in the air, so both sample and lenses have to be kept in high vacuum conditions. Both vacuum conditions and irradiation with an accelerated electron beam are damaging to the biological sample. To overcome this problem, several specific sample preparation techniques have been devised. In the field of macromolecular 3DEM, the most used are: negative staining, and vitrification. In negative staining, the biological sample is incubated with a high weight metal salt that forms a layer leaving vacant the areas where protein is present, creating a sort of mask around the specimen. The preparation is dried before imaging. In vitrification, the sample is frozen by fast immersion in liquid ethane, kept and imaged by cryoEM at liquid nitrogen temperatures, without drying. This preserves the sample in a state closer to physiological, as opposed to negative staining which may alter structural integrity by interaction with the staining agent, flattening and drying. However, cryoEM requires more expertise for working at low temperatures, provides less contrast and requires higher sample concentration than negative staining. Negative staining is essential for initial characterization of new specimens, as well as for full analysis of low abundance or small complexes, which would be difficult to pick in cryoEM images.

Although in principle images taken with a TEM can reach atomic resolution or better, for biological specimens several factors limit the resolution attained. In cryo-EM, the resolution is mainly limited by image quality (aberrations and noise), while in negative staining the resolution is limited by the penetration capability of the staining agent (grain size). A critical factor to limit resolution in both techniques is sample homogeneity, since 3DEM reconstruction algorithms require averaging of many different images which are assumed to be originated from the same object, as will be explained later on. In the most favorable cases, with highly homogeneous and symmetric specimens, near atomic resolution has been reached. This was the case, for example, for two-dimensional crystals of bacteriorhodopsin (~3.6Å) (Subramaniam and Henderson, 2000), the cytoplasmic polyhedrosis virus (CPV) (~3.88 Å) (Yu, Jin and Zhou, 2008), or helical filaments of F-actin (10 nm in diameter) (Walther, 2008), human adenovirus type 5(Liu, Jin, Koh, Atanasov, Schein, Wu and Zhou, 2010). Near atomic resolution in helical filaments has been reached for things like microtubules, flagella, or acetylcholine receptor (Yonekura, Maki-Yonekura and Namba, 2003). Even at lower resolutions, valuable information regarding complex structure, assembly and function can be obtained, particularly if high resolution structures of individual proteins or domains are available and can be fitted into a 3DEM map of the whole complex.

The most unfavorable case for 3DEM are small complexes (MW < 200 kDa) without symmetry and roughly globular in shape; for these specimens it is much harder to recognize the particles of interest in the micrograph, and image processing algorithms may fail due to the lack of enough signal in the noisy projections. The problem is even worse if such a complex turns out to be heterogeneous in composition or structure; however, structural heterogeneity may reflect the natural flexibility of the complex and be required for its function, and so study of these “hard” specimens is of biological relevance. Problems derived from sample heterogeneity can be alleviated either by chemical (cross-linking) or computational (image classification) methods (Leschziner and Nogales, 2007; Spahn and Penczek, 2009).

1.6 3DEM: the basics

Thanks to the depth of focus of a TEM, the 2D images obtained for a typical macromolecular complex in biology can be considered true projections of the 3D specimen. The process that allows obtaining 3D information from 2D projections is named three-dimensional reconstruction. The basis of 3D reconstruction is the *central section theorem*, which states that the Fourier transform of the (n-1) dimensional projection of the n-dimensional object is identical to a central (n-1)-dimensional plane of the n-dimensional Fourier transform of the object (Frank, 2006).

Ideally, to obtain a 3D map of an object one should record 2D projections of the object in all possible orientations, to have enough 2D sections to fill the 3D Fourier transform of the object. However, in 3DEM only a limited set of projections from the same object can be recorded. There are two reasons for this limitation: the physical impossibility to rotate the sample holder to angles higher than 60-70º, and the radiation damage suffered by the sample, which undergoes structural alterations after being irradiated with a few tenths of electrons. To overcome this last problem, we can take advantage of the fact that our grid contains a large number of copies of the same macromolecule, lying (in principle) in random orientations on the support. So, if we are able to determine their relative orientations, we can combine these projections of individual particles to obtain an average 3D map. When using this method, we are assuming that all 2D projections in our dataset are originated by the same object. If this is not the case, the 3D reconstruction obtained will not be a faithful representation of our specimen. This is why, apart from robust alignment algorithms, 3DEM has to rely on powerful image classification techniques, particularly in the case of flexible biological complexes.

A general workflow for image processing in 3DEM is shown in Figure 8. Different software packages can be used to perform each step, depending on their particular strengths or specialized algorithms.

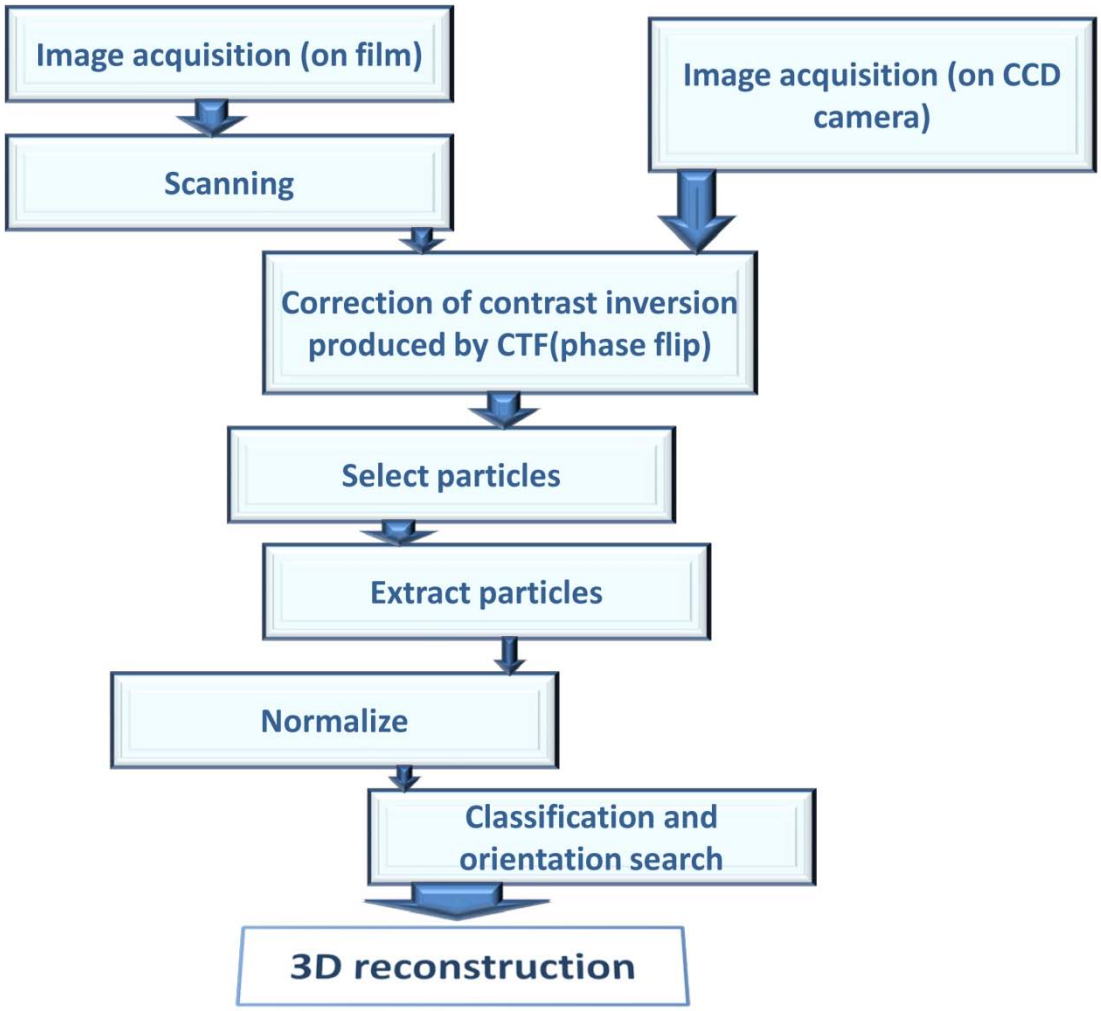


Figure 8. General 3D reconstruction protocol.

1.7 3DEM: the importance of an initial model

Orientation search is a crucial step in 3DEM. All particle projections must be brought to a common origin and rotation reference system, so that they can be accurately merged into a 3D reconstruction. The most successful applications in the field use the method called “projection matching” (Penczek, Grassucci and Frank, 1994). This method consists in comparing our experimental images with projections calculated from a 3D template. Using comparison criteria such as cross-correlation, each image is assigned the orientation parameters corresponding to the best matching template projection. A new 3D reconstruction is calculated using the obtained shifts and angles with the experimental data, and used as template for a new projection matching round. This procedure is iterated until convergence, that is, until the

orientation parameters or the new 3D map are indistinguishable from those obtained in the previous iteration.

Projection matching has been very successful for refining datasets for which a previous low resolution 3D map was available (Glaeser, 2008). However, for determining a 3DEM map *de novo*, one has to be particularly careful, because projection matching is very sensitive to bias by the initial model. If the latter is not compatible with our data, either because it is too different from the object under investigation, or because the dataset itself is not consistent, one can end up with a non-converging refinement, or worse, with a completely artifactual 3D map resembling our initial model but not our specimen (Grigorieff, 2000; Penczek, 2002; Shaikh, Hegerl and Frank, 2003).

On starting a new 3DEM project, there are several ways to obtain an initial model for our refinements. One possibility is using a synthetic 3D model, built from geometrical shapes, according to our expectation of how our specimen will look like (e.g. based on the size and shape of the 2D projections). If the structure of a homologous complex is available, this can also be used as an initial model, after filtering to low resolution. An initial 3D model can also be calculated from the experimental images, by using a common lines approach (Crowther, 1970). Finally, an initial model can be calculated by the *random conical tilt* (RCT) approach (Radermacher, 1988).

In the RCT approach, the same area in the sample is photographed tilted at 55º, then untilted (Figure 9). The untilted dataset is aligned and classified in 2D; then the aligment parameters found, together with the tilt geometry, are applied to the corresponding tilted particles to calculate the 3D map. If different classes are detected in the untilted images, each class average can be separately reconstructed. This produces several 3D maps that can be compared, aligned in 3D and merged to produce an average 3D structure, as long as it can be determined that they arise from the same object observed from different points of view.

In general, the safest procedure to ensure unbiased structure determination by 3DEM is to refine the same dataset starting from different initial references, and see if

all refinements converge to similar solutions. If not, sample or dataset heterogeneity must be reassessed.

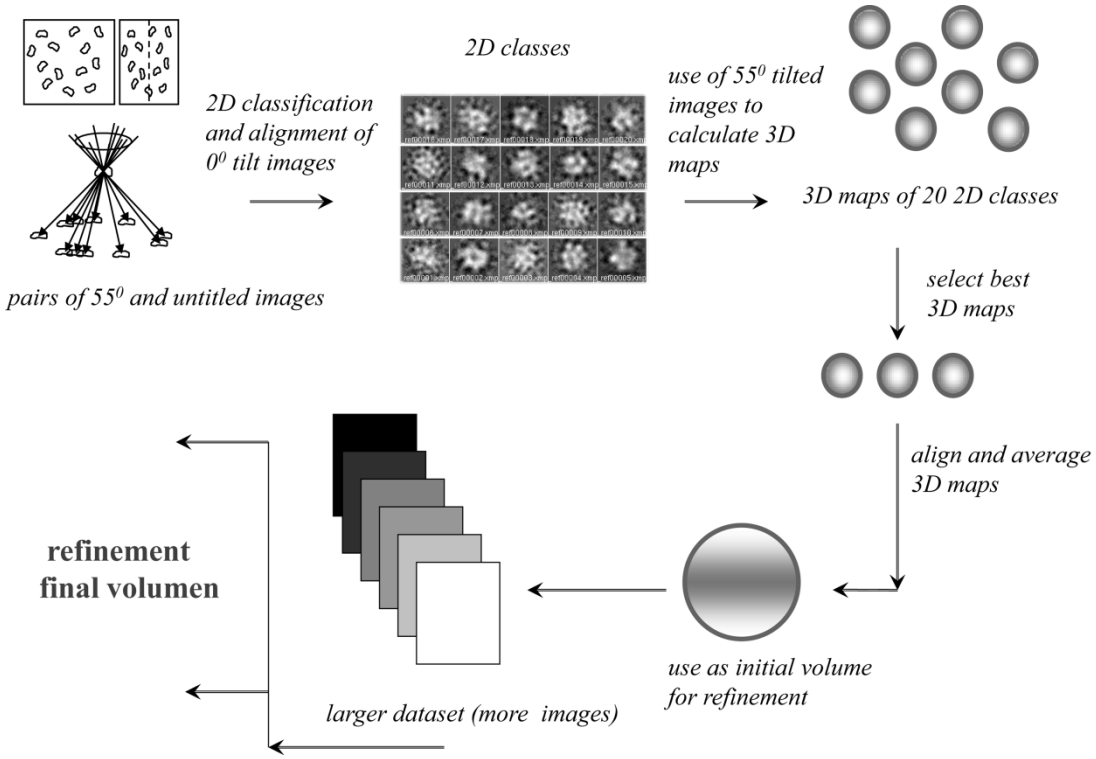
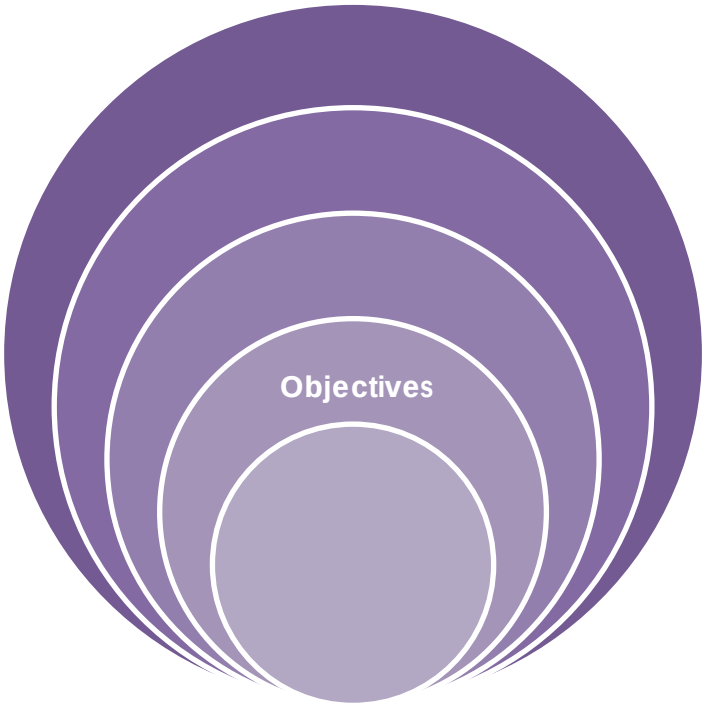


Figure 9. Scheme of work using the RCT method.


Objectives

2. *Objectives*

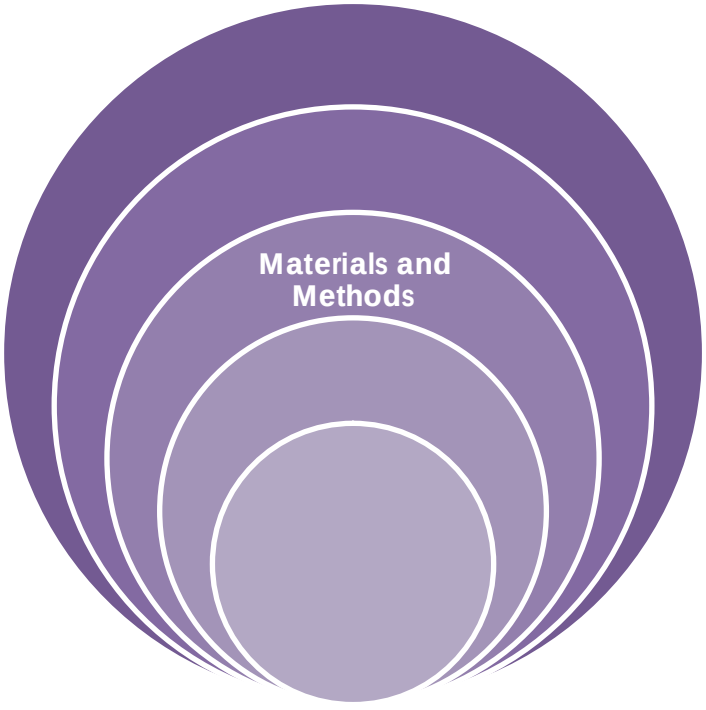
The general aim of this thesis is to obtain new data on the structure of different DNA pol III clamp loader subcomplexes, which could help to understand the dynamic assembly of the enzyme. In particular, the following objectives have been pursued:

- Obtain 3DEM maps of the *E. coli* clamp loader subcomplexes γ_4 and $\gamma_4\psi\chi$.

- Determine the changes in protein-protein contacts between γ molecules in γ_4 (our EM analysis) and $\gamma_3\delta\delta`$ (minimum clamp loader crystal structure) to explain how γ changes between tetramer and trimer depending on the presence of other clamp loader subunits (δ and δ`).

- Determine the organization of the C-terminal region of γ, absent in the crystal structure, by difference mapping between the homotetramers of full length and truncated γ.

- Determine the localization of the ψχ heterodimer by difference mapping between γ_4 and $\gamma_4\psi\chi$ 3DEM maps.


Materials and
Methods

3. *Materials and methods*

3.1 Proteins

Protein complexes analyzed in this thesis were provided by Prof. Charles S. McHenry (Department of Chemistry and Biochemistry, University of Colorado at Boulder, Colorado, USA), and are listed in Table 3. .

Table 3. Complexes analyzed in this thesis.

Protein complex		**Molecular weight of the complex (kDa)**	**Concentration (mg/ml)**	**Buffer composition**
E. coli	$\gamma(370)_4$	165, 000	17	25 mM HEPES pH 7.5, 25 mM NaCl,0.1 mM EDTA
	$\gamma(370)_4\psi\chi$	197,000	45	25 mM HEPES pH 7.5, ~300 mM NaCl, 0.1 mM EDTA
	$\gamma(430)_4$	190,000	18.7	25 mM HEPES pH 7.5, 25 mM NaCl,0.1 mM EDTA
	$\gamma(430)_4\psi\chi$	222,000	36.0	25 mM HEPES pH 7.5, ~250 mM NaCl, 0.1 mM EDTA

Briefly, proteins were expressed in *E. coli* and purified by several chromatography steps (Table 4). Both full length and a truncated form of γ (containing residues 1-370) were purified as tetramers. ψ and χ were expressed together in an artificial operon and purified as a heterodimer. To form the $\gamma_4\psi\chi$ complexes, γ subunit (10 nmol of monomer) was mixed with an excess of the ψχ dimer (20 nmol of each subunit monomer) in a volume of 0.5 ml. Proteins were incubated at room temperature for 15 min before injection into a Superose 6 column, equilibrated in 20 mM sodium phosphate (pH 7.4), 10% glycerol, 50 mM NaCl, and 5 mM $MgCl_2$. Collected fractions were then injected into a *Mono Q Ion Exchange* column, and eluted with a 15-ml linear gradient from 50 to 400 mM NaCl in the same

buffer. All subunits and complexes eluted at unique conductivities in the following order: $\chi\psi$ (80 mM NaCl), γ_4 (260 mM NaCl) and $\gamma_4\chi\psi$ (280 mM NaCl).

Table 4. Short scheme of purification of subcomplexes.

Subunit	1st step	2nd step	3rd step	Reference
≻	Q-Sepharose Chromatography 50-400mM Tris-HCl (pH 7.5), 50 mM NaCl, 10% (w/v) glycerol, 5 mM DTT)	Sephacryl S-300 HR Gel Filtration Chromatography 25mM Hepes-KOH (pH 7.5), 25 mM NaCl, 5% glycerol, 0.1 mM EDTA		(Dallmann, Thimmig and McHenry, 1995)
ψχ	Q-Sepharose Chromatography 20% glycerol, 50 mM Tris-HCl (pH 7.8), 1 mM EDTA, 5 mM DTT 20-200mM NaCl	SP-Sepharose Chromatography 20% glycerol, 50 mM Tris-HCl (pH 7.0), 1 mM EDTA, 5 mM DTT 20-200mM NaCl	Sephacryl S-100 Gel Filtration Chromatography 20% glycerol, 20 mM potassium phosphate (pH 7.5), 0.5 mM EDTA, 20 mM NaCl, 2.5 mM DTT	(Olson, et al., 1995)

All provided complexes were analyzed by SDS-PAGE gel electrophoresis and gel filtration techniques to check for purity and oligomerization state before using them for EM preparations.

3.2 Glutaraldehyde cross-linking and gel filtration

Fresh SIGMA EM grade 25% glutaraldehyde stock was first diluted in water, then adjusted to the final concentration in the protein solution. The reaction mixture was incubated for 10 min at 25 ^{0}C in a thermo-mixer with slow mixing rate. To stop the reaction, either SDS protein electrophoresis loading buffer, or (in case of gel filtration experiments) 2M NH_4Cl was added to its final concentration of 200mM and incubated for 10 min at room temperature. The pretreated sample was injected into a Superdex 200 10/30 column, calibrated with the Amersham kit standards: ferritin (440 kDa), catalase (222 kDa), aldolase (158 kDa), albumin (67 kDa), ovalbumin (43 kDa).Column output was analyzed by absorbance (OD_{280nm}).

3.3 GraFix

GraFix is a method of sample preparation for 3DEM which combines sedimentation by centrifugation in a glycerol gradient with mild chemical fixation (Figure 10). The idea is to achieve both structural preservation and homogeneity, eluding the risk of aggregation, for flexible or labile macromolecular complexes (Kastner, Fischer, Golas, Sander, Dube, Boehringer, Hartmuth, Deckert, Hauer, Wolf, Uchtenhagen, Urlaub, Herzog, Peters, Poerschke, Luhrmann and Stark, 2008).

For each complex to be analyzed, a volume of 300 µl at a concentration of 150 µg/ml was deposited on top of a 5%-30% glycerol gradient in buffer corresponding to each complex. The gradients were formed using a Gradient Master device (BioComp), that tilts the tubes containing the heavy and light solutions to the proper angle (depending on each gradient), rotates them for a specific time and then returns them to their original vertical position, to produce a continuous gradient in a reproducible way. The heavy solution contained 0.15% of glutaraldehyde. Ultracentrifugation was performed at 25.000 rpm for 18 hours at 4ºC using an SW 55 Ti rotor in a Beckman XL-90 ultracentrifuge. After centrifugation, fractions were collected from the top using a BioComp Piston Gradient Fractionator and analyzed by SDS PAGE and EM. As a control, a gradient without gluteraldehyde was run in parallel.

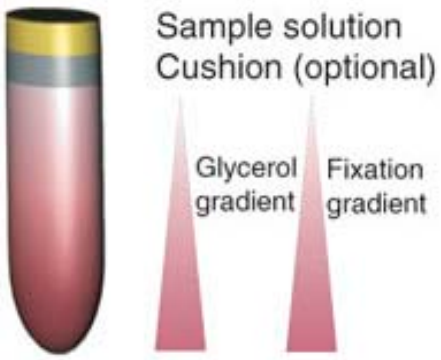


Figure 10. Schematics of the GraFix method. Glutaraldehyde is added to the most concentrated glycerol solution. The sample is added on top of the tube. Modified from (Kastner, et al., 2008).

3.4 Protein electrophoresis

Samples were analyzed by electrophoresis in either 10% or 4-20% polyacrylamide gels without previous boiling of the loaded sample, only incubation for 5 min at 86^0C in case of cross-linked samples, and 99^0 in case of normal samples. Gels were dyed either with Coomassie blue using standard protocols, or silver with the modifications indicated in Table 5. To avoid staining artifacts caused by the presence of glutaraldehyde in the samples, SYPRO Ruby Protein Gel Stain (Sigma-Aldrich S5692) was used. SYPRO Ruby stained gels were scanned at 532 nm in a Typhoon laser scanner or GelDoc Imaging System (BioRad).

Table 5. Silver staining for proteins. The standard protocol of Amersham Bioscience was modified as follows.

Step	**Solutions**	**Time**
Fixation	Ethanol 100 ml Acetic acid 25 ml up to 250 ml	15 min
Sensitizing	Ethanol 75 ml Sodium thiosulphate (5% w/v) 10 ml Sodium acetate (17g) up to 250 ml	15 min
Washing	Distilled water	3x3min
Silver reaction	Silver nitrate solution (2.5% w/v) 2 ml up to 30 ml	10 min
Washing	Distilled water	2x 30 sec
Developing	Sodium carbonate (6.25g) Formaldehyde 50 µl up to 250 ml	2-5 min
Stopping	EDTA-Na2*2H2O (3.65g) up to 250 ml	5 min
Washing	Distilled water	2x5 min

3.5 Preparation of samples for electron microscopy

For negative staining, 10 µl of protein sample at the appropriate concentration was adsorbed for 5 min onto a glow-discharged, collodion/carbon coated copper grid. Then the grid was washed with 2 drops of 2% (w/v) uranyl acetate and incubated for 1

min on 10 µl of the same staining agent (Figure 11). Washing with staining agent instead of buffer or water improved EM preparations by producing a more homogeneous staining (Miyata, Oyama, Mayanagi, Ishino, Ishino and Morikawa, 2004).

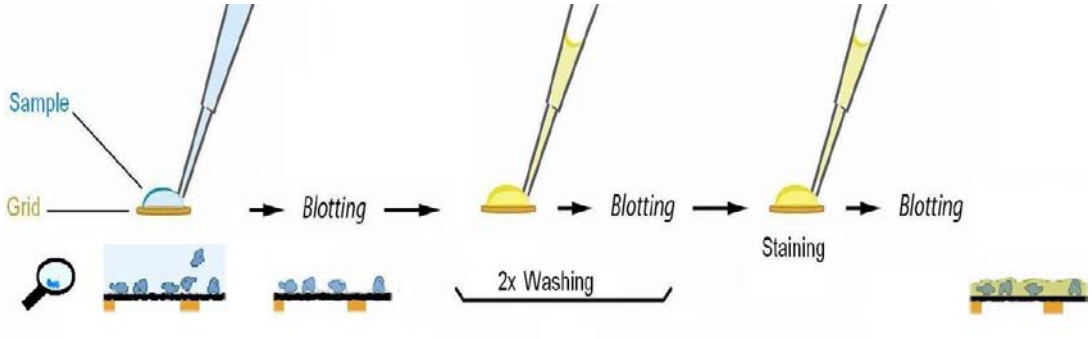


Figure 11. Negative staining procedure. The sample is adsorbed to the carbon support for ~3min. Then the rest of the protein not adsorbed is removed by blotting. The grid is washed (2x~30sec) with uranyl acetate and finally stained for one additional minute. Drawing adapted from (Gómez-Llorente, 2007).

For cryoEM, samples prepared by the GraFix method were spun in a Microcon YM-100 centrifugation device to increase protein concentration and to exchange into a glycerol-free buffer. A thin layer of mica-deposited carbon was floated onto Quantifoil R2/4 holey carbon grids. After drying, grids were glow discharged and incubated with the sample for 1 min before manual blotting and vitrification in liquid ethane using a Leica CPC plunger. Grids were stored in liquid nitrogen until examined in the microscope.

3.6 Image acquisition

Grids were examined in either a Tecnai F20 transmission electron microscope at 200 kV (for RCT reconstruction and cryoEM), or at 100 kV in a JEOL JEM 1230. All micrographs were taken at low electron dose conditions. CryoEM images were recorded at 67600x magnification (with defocus values from -3 to -8 µm) using a 16 Mpx FEI Eagle CCD camera with a 15 µm pixel size. Negative staining micrographs were recorded on film at 60000x magnification (62000x for RCT) and digitized with a Zeiss SCAI scanner at 7 µm pixel size.

3.7 Image preprocessing

All image processing was carried out using the XMIPP Software Package (Scheres, Nuñez-Ramirez, Sorzano, Carazo and Marabini, 2008), except where indicated. To reduce computational demands, scanned micrographs were downsampled by a factor of 2 giving a pixel size of 2.26 Å for RCT data, 2.22 Å for cryoEM (no downsampling), and 2.33 Å for all other datasets. Micrograph Fourier transforms were calculated by patch averaging, and CTF parameters were estimated, using CTFFIND3 (Mindell and Grigorieff, 2003). Micrographs presenting minimal drift and astigmatism were selected for further processing. Contrast reversion due to the CTF oscillation was corrected in the micrographs ("phase flipping").but in case of RCT data. Particles were manually selected, extracted from the micrographs in 80x80 px frames and normalized, then scaled to 40x40 pixels (no scaling for RCT). This gave a final pixel size of 4.66 Å for negative staining and 4.44 Å for cryoEM images. "Junk" (outlier) particles were removed after dataset sorting according to basic image statistics using *xmipp_sort_by_statistics*.

Reference free, simultaneous 2D alignment and classification of images was performed using maximum-likelihood based algorithms in Fourier space (MLF2D) (Scheres, Nuñez-Ramirez, Gomez-Llorente, San Martin, Eggermont and Carazo, 2007). This approach allowed us to introduce a high frequency limitation in the probability calculations, which was critical to reduce artifacts derived from background noise alignment in our small specimen, low signal images. Images corresponding to class averages clearly representing noise, or too close neighboring particles, were discarded at this point.

3.8 3D reconstruction

Related crystallographic structures, as well as RCT reconstructions calculated from our EM data were tested as initial references for orientation search.

- Crystallographic structures: the structure of the minimal clamp loader (PDB ID: 1jr3, (Jeruzalmi, et al., 2001)), containing three γ(370) molecules plus δ

and δ', was converted into an EM-like density map using EMAN program *pdb2mrc*, with the pixel size of our EM datasets, and low-pass filtered to 40 Å resolution (Figure 12). This model is referred to as "1JR3" throughout this thesis. An alternative model was constructed keeping only the three γ(370) molecules. This will be referred to as "PDB".

- RCT reconstruction: micrograph pairs were recorded at 55° and 0° tilt angles. Pairs of individual particles were picked, extracted and normalized. The 0º tilt dataset was aligned and classified into 20 classes using *XMIPP M*LF2D, with the high frequency cutoff set at 20 Å. The 2D alignment parameters thus obtained were applied to the 55^0 tilt images, together with the geometrical tilt parameters, to calculate a 3D map from each class. Selected 3D maps (see *Results*) were aligned and averaged to produce the initial model referred to as "RCT" throughout this thesis.

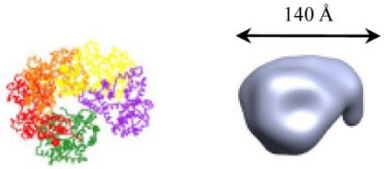
Model	**1JR3**	**PDB**
Molecules	$\gamma(370)_3\delta\delta'$	$\gamma(370)_3$
C-terminal view		
N-terminal view		
Side view		

Figure 12. Density maps generated from the clamp loader crystal structure, complete (1JR3) or including only three γ monomers (PDB). The Cα trace in the crystal structures is colored by molecule. The 40 Å filtered density is shown in cyan.

Orientation refinement and 3D reconstruction were performed either using the standard EMAN refinement procedure (Ludtke, Baldwin and Chiu, 1999), cross-correlation based projection matching in XMIPP, or Fourier space maximum-likelihood refinement (MLF3D) (Scheres, et al., 2007). The best results were obtained using MLF3D, with 15° angular sampling and high frequency cutoff at 35 Å. All initial 3D references for MLF3D were low pass filtered to 80 Å resolution to reduce model bias.

3.9 3D map interpretation

UCSF CHIMERA (Pettersen, Goddard, Huang, Couch, Greenblatt, Meng and Ferrin, 2004) was used to visualize 3D maps. A contour threshold corresponding to 80-100 % (for a good comparison of different 3D maps) of the expected volume for each protein complex was used. This was calculated by taking into account the average protein density 1.33 g/cm^3. Maps were manually aligned in 3D, then the alignment was optimized using CHIMERA *FitMap* tool. For difference mapping, the gray levels in the 3D maps to be subtracted were normalized within a common region, defined by a mask loosely enclosing 100% of the expected volume for the complexes.

Crystal structures were fitted into the 3D maps with UROX (Siebert and Navaza, 2009). UROX can refine the positions of several molecules within a map simultaneously, considering each molecule as an independent rigid body.


Results

4. *Results*

4.1 Initial sample characterization

Samples of $\gamma(370)_4$, $\gamma(370)_4\psi\chi$, $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$ were characterized by electrophoresis and gel filtration. SDS-PAGE indicated that all complexes were highly pure and free of degradation (Figure 13).

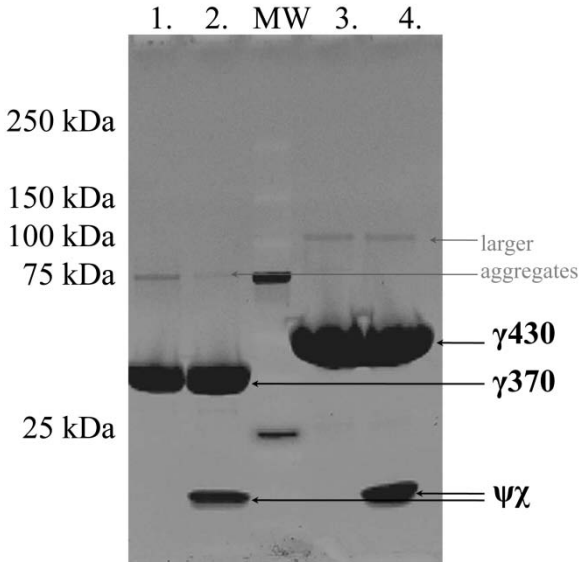


Figure 13. Electrophoresis characterization of complexes. 4-15% acrylamide gradient SDS gel stained with SYPRO Ruby. 1. $\gamma(370)_4$; 2. $\gamma(370)_4\psi\chi$; 3. $\gamma(430)_4$; 4. $\gamma(430)_4\psi\chi$. ψ and χ are not resolved in this kind of gel.

In size exclusion chromatography analysis, all complexes showed a majoritary peak consistent with the expected molecular weight (Figures 15, 16a, and 17b). A small amount of larger aggregates was observed in some cases. Both SDS-PAGE and gel filtration experiments indicated biochemical homogeneity of the samples. However, imaging by negative staining at the EM revealed a heterogeneous population of particles, with different sizes and shapes (Figure 14). Attempts at image processing of this kind of preparations did not converge to a single solution.

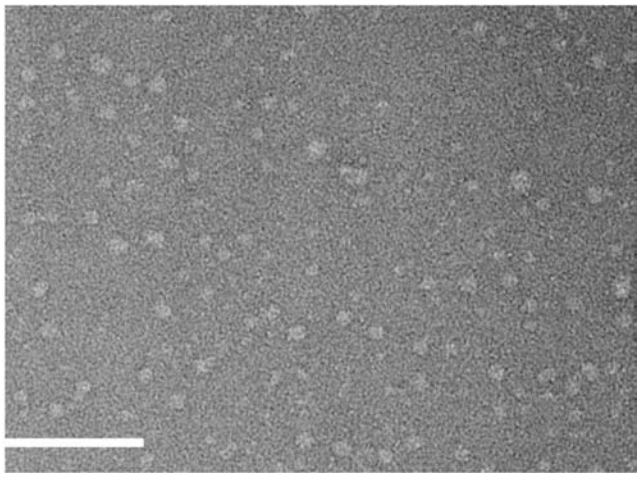


Figure 14. Micrograph of $\gamma(430)_4\psi\chi$ showing heterogeneity in particle size and shape. Scale bar stands for 100 nm .

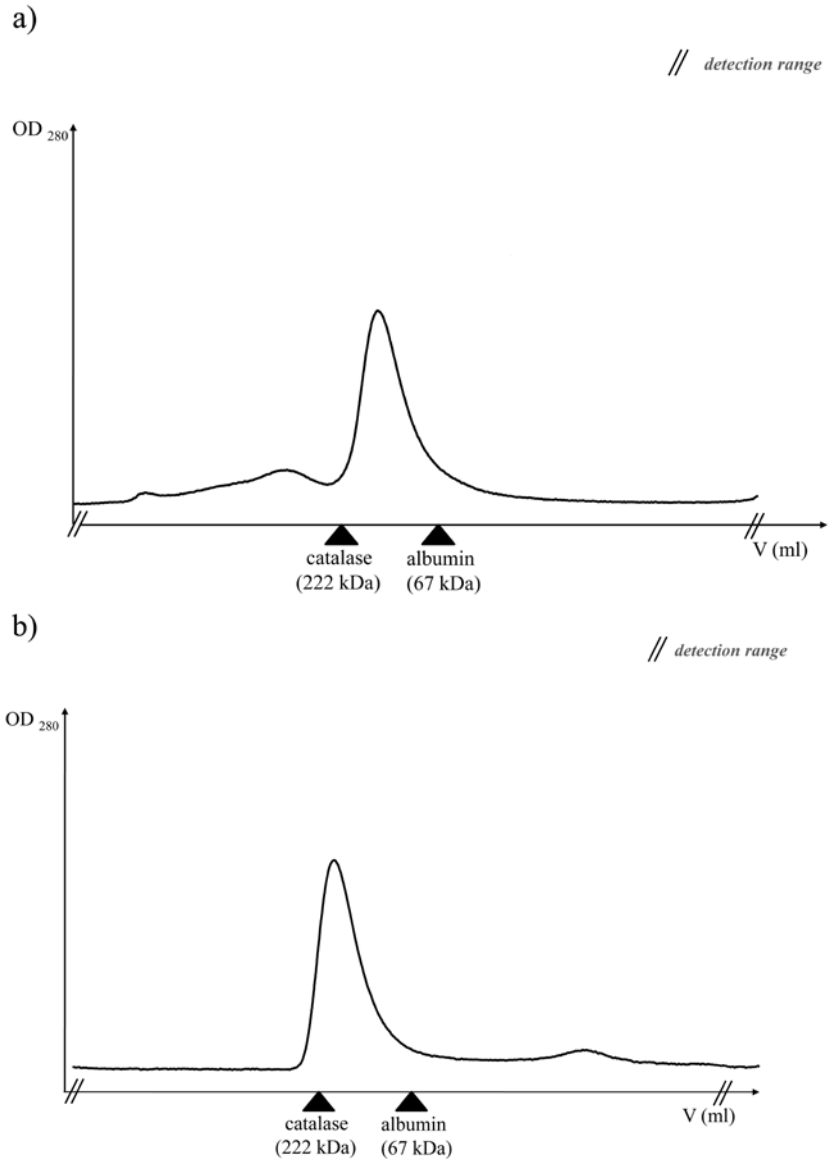


Figure 15 Oligomerization analysis of(a) $\gamma(370)_4$ and (b) $\gamma(370)_4\chi\psi$ tetramers using Superdex 200 10/30 gel filtration chromatography. Elution position of molecular weights standards of 222 kDa and 67 kDa is shown.

4.2 Cross-linking and gel filtration

Since the heterogeneity observed at the EM could be caused by interaction of the complexes with the staining agent, we decided to investigate whether mild cross-linking previous to sample preparation for EM would improve the structural preservation of the complexes.

To define cross-linking conditions, different gluteraldehyde concentrations were tested (Figure 16). A final concentration was chosen such that some non cross-linked protein was still observed in denaturing gel electrophoresis. This election was based on the consideration that stronger cross-linking conditions might give rise to unspecific protein aggregation. The gluteraldehyde concentration chosen was 0.035%.

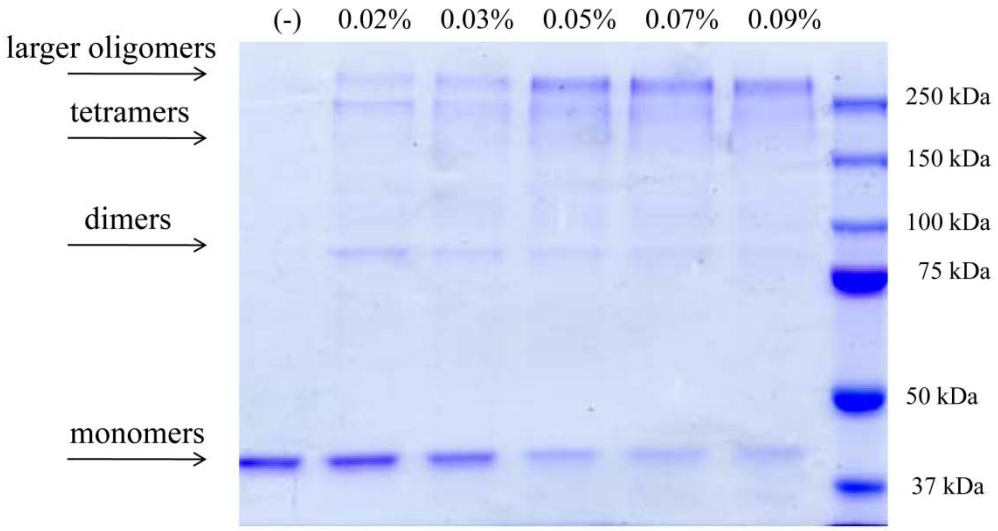


Figure 16. 10% polyacrylamide gel presenting cross-linking of $\gamma(370)_4$ with different final concentrations of glutaraldehyde.

Next, samples of the full length γ protein complexes $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$ were cross-linked under the condition selected and subject to gel filtration. Fractions corresponding to the majoritary absorbance peak were collected and analyzed by SDS-PAGE (Figure 17b) and negative staining EM. Gel filtration indicated that cross-linking did not alter the oligomeric state of the $\gamma(430)_4$ complex, which presented esentially the same elution profile both with and without gluteraldehyde incubation (Figure 17a). The effect of cross-linking resulted in a marked improvement in the

structural homogeneity of the sample, as judged by negative staining EM (Figure 17c).

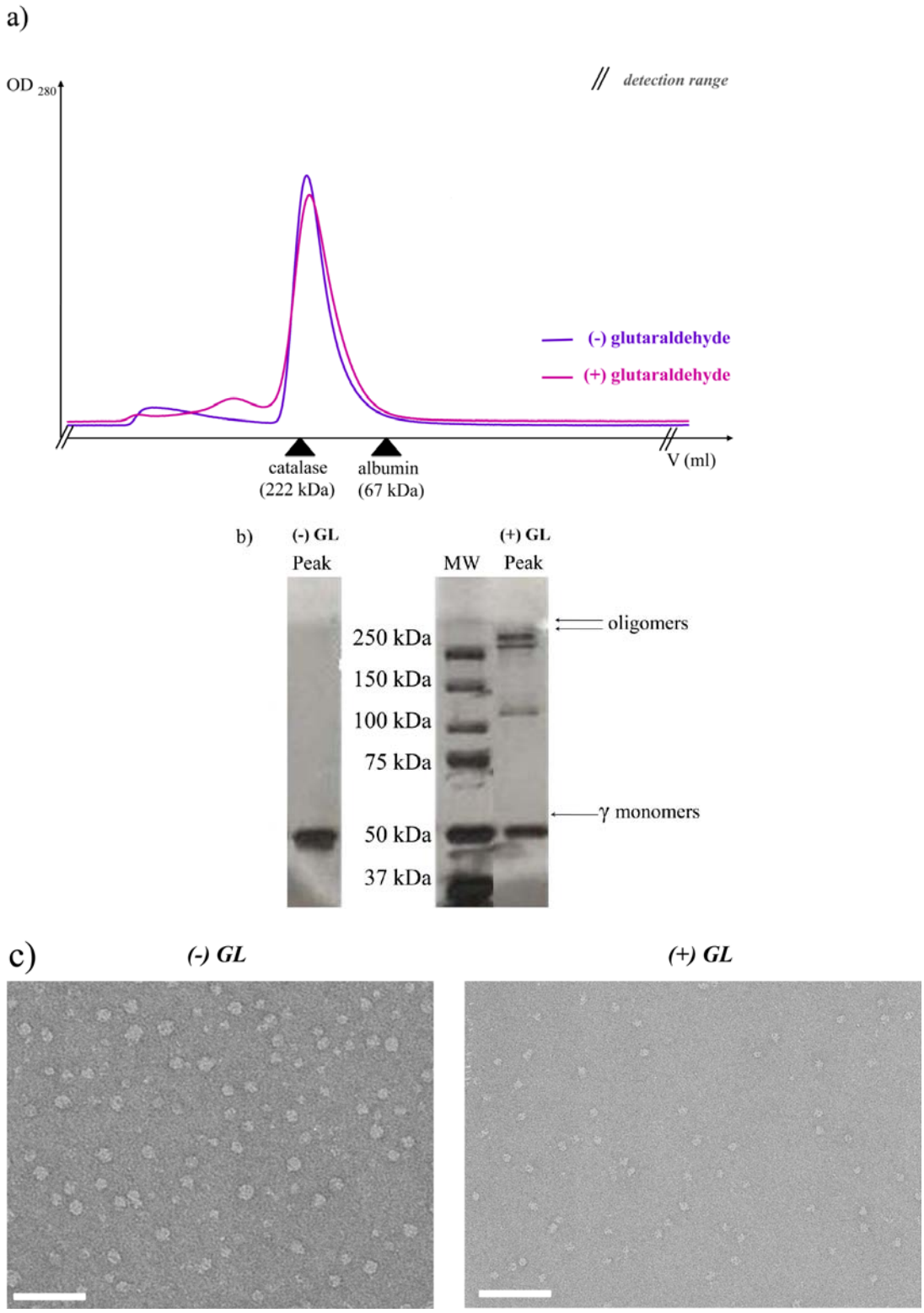


Figure 17. (a) Oligomerization analysis of $\gamma(430)_4$ with and without cross-linking with 0.035% glutaraldehyde (standards elution volumes and MW are shown). (b) 10% SDS polyacrylamide gel analysis of $\gamma(430)_4$ elution peak (the gel was stained with silver protein dye) .(c) Micrographs showing heterogeneity reduction in the presence or gluteraldehyde. Scale bar stands for 100 nm.

For the $\gamma(430)_4\psi\chi$complex, cross-linking induced moderate aggregation (Figure 18b 18a). Particle homogeneity at the EM was improved, although not as markedly as for the $\gamma(430)_4$ homotetramer (Figure 18c).

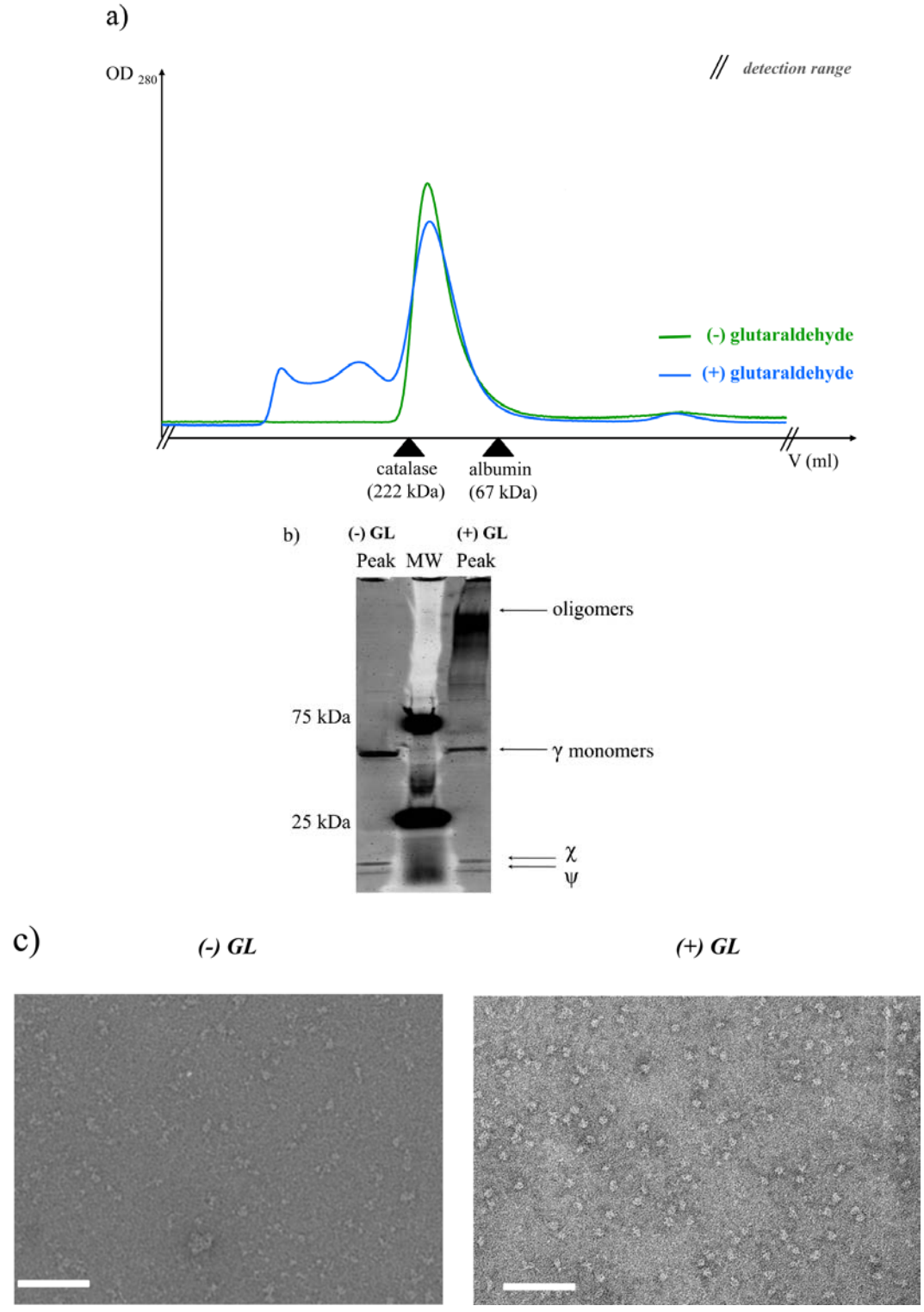


Figure 18. (a) Oligomerization analysis of $\gamma(430)_4\psi\chi$ with and without cross-linking with 0.035% glutaraldehyde (standards elution volumes and MW are shown) (b) 10% SDS polyacrylamide gel analysis of $\gamma(430)_4$ $\psi\chi$ elution peak (the gel was stained with silver protein dye) (c) Micrographs showing heterogeneity reduction in the presence or gluteraldehyde. Scale bar stands for 100 nm.

4.3 Calculation of RCT models

To obtain initial 3D models for the $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$ complexes, the random conical tilt method was used with samples prepared by cross-linking and gel filtration (Figure 19). Micrographs were taken at 55 and 0º tilt (46 micrograph pairs for $\gamma(430)_4$; 45 micrograph pair for $\gamma(430)_4\psi\chi$), and from these 8800 particle pairs were picked for $\gamma(430)_4$, 8200 for $\gamma(430)_4\psi\chi$.

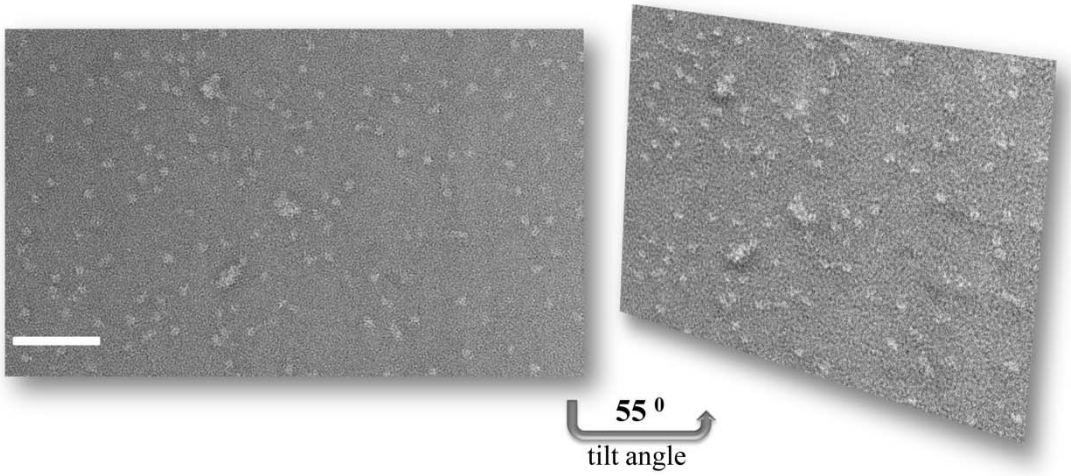


Figure 19. Example of a micrograph pair of $\gamma(430)_4$ cross-linked with 0.035% glutaraldehyde, taken at 0^0 and 55^0 tilt angle. Scale bar stands for 100 nm.

Images taken at 0º tilt were aligned and classified into 20 groups using XMIPP MLF2D, with the high frequency cutoff set at 40 Å (Figure 20). No preferential view was detected. A 3D map was calculated for each MLF2D group using the alignment parameters obtained and the corresponding 55º tilted images, and filtered down to 40 Å resolution. It was not straightforward to decide if all maps were representations of the same object observed from different points of view or if they were originated by different objects, particularly in the case of $\gamma(430)_4\psi\chi$. Therefore, it was decided to obtain an average RCT model for each complex using only those single 3D maps that fulfilled certain quality criteria (Figure 21).

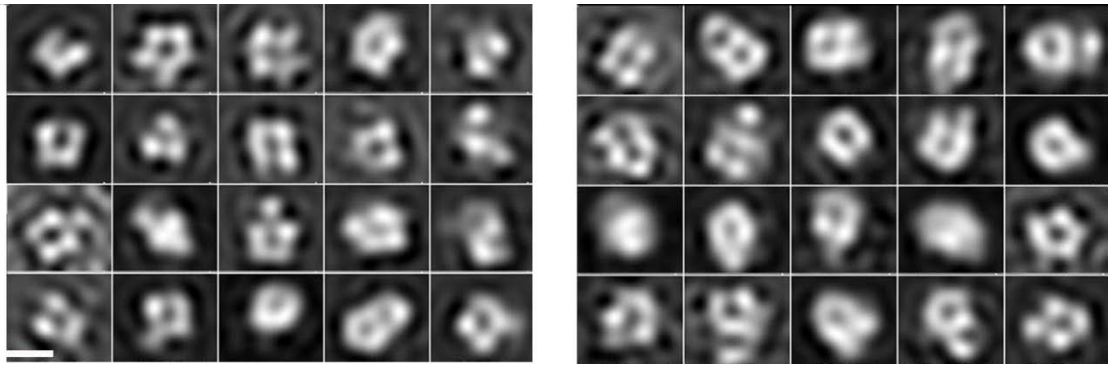


Figure 20. MLF2D classification of $\gamma(430)_4$ (left-hand side) and $\gamma(430)_4\psi\chi$ (right hand side) untilted data. Scale bar stands for 10 nm.

First, low populated groups producing very noisy 3D maps were discarded. This resulted in elimination of 6 groups with less than 200 particles in $\gamma(430)_4$, and 6 groups with less than 150 particles in the $\gamma(430)_4\psi\chi$ dataset. The remaining maps were aligned in 3D, first manually and then using USCF CHIMERA *FitMap*, using one of them as a reference (maps in orange and green circle in Figure 21). The reference was chosen due to the clearance of the 3D map and higher number of particles that went to this particular class. The cross-correlation coefficients given by CHIMERA *FitMap* were used as a criterion to further discard the worse-fitting groups (2 groups in $\gamma(430)_4$, 4 groups in $\gamma(430)_4\psi\chi$). The aligned, remaining 3D maps for each complex were averaged to produce the RCT maps that would serve as initial models in further image processing refinements. These final maps included data from 5200 particles for $\gamma(430)_4$, and 4000 particles for $\gamma(430)_4\psi\chi$.

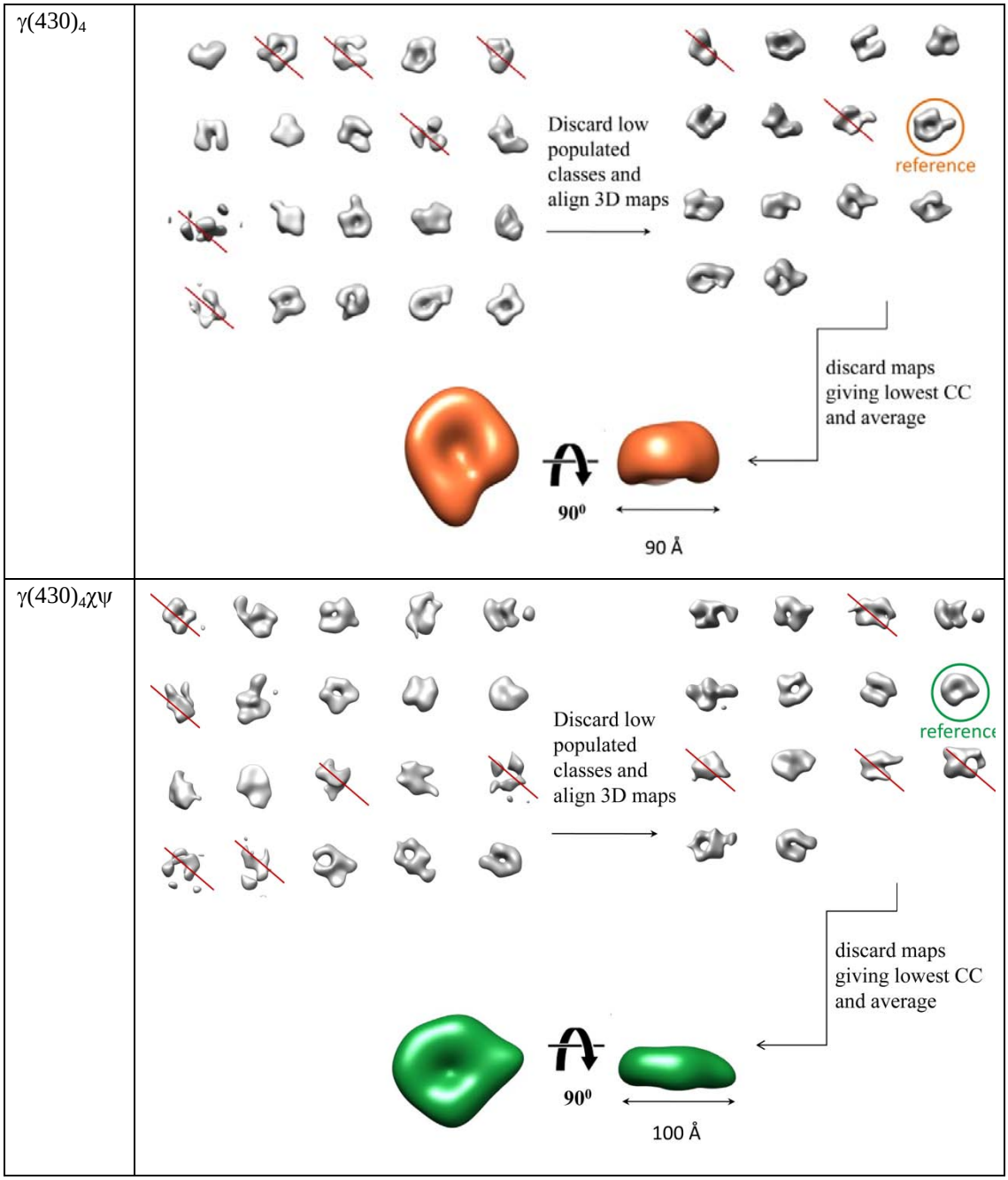


Figure 21. Scheme for obtaining $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$ average RCT reconstructions.

After calculating the RCT maps, additional image datasets of 25000 particles for $\gamma(430)_4$ and 28000 particles for $\gamma(430)_4\psi\chi$ were acquired, preprocessed and used for orientation search and 3D reconstruction. Several initial models were tested to assess convergence of the refinements, among them either RCT maps or PDB (see Section 3.8). Different refinement strategies were tried, both in XMIPP and in EMAN: using soft masks to reduce the influence of background noise; using very stringent criteria to discard inconsistent particles; filtering the input images and reference projections to different resolutions; etc. Unfortunately none of the procedures resulted in

convergence, as different final maps were obtained from the same dataset depending on the initial model. Both from the RCT and these latest results, we conclude that even after cross-linking and gel filtration, a considerable degree of structural heterogeneity was still present in the samples.

4.4 GraFix

A new attempt at reducing structural heterogeneity was carried out using a different cross-linking method: GraFix (Kastner, et al., 2008). We reasoned that the combination of gluteraldehyde and glycerol gradient might allow stronger cross-linking conditions (larger gluteraldehyde concentrations) without producing large aggregates. All four protein complexes ($\gamma(370)_4$, $\gamma(370)_4\psi\chi$, $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$) were therefore subject to sedimentation in a glycerol/gluteraldehyde gradient as described in Section 3.3. All collected fractions were analyzed in SDS 4-20% gels (Figure 22). Cross-linked complexes were observed between 1/3 and 2/3 the height of the tube, that corresponds to glutaraldehyde concentrations between 0.04-0.06% .

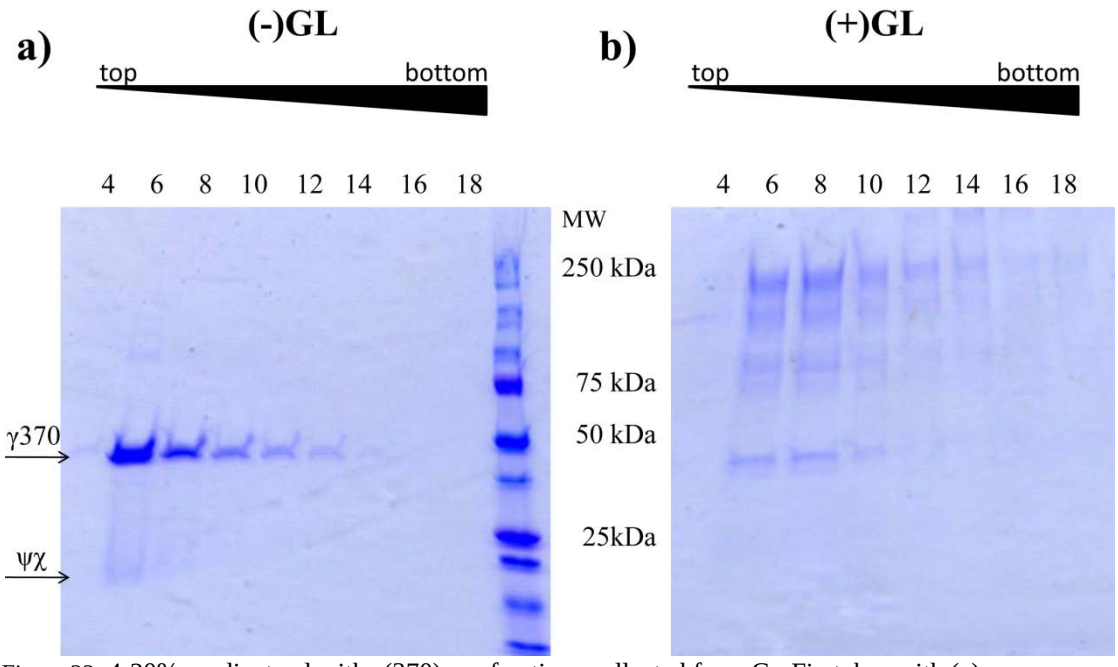


Figure 22. 4-20% gradient gel with $\gamma(370)_4\psi\chi$ fractions collected from GraFix tubes with (a) no glutaraldehyde and (b) 0-0.15% gradient of glutaraldehyde.(up to 25 fractions were collected, data not demonstrated).

Fractions that corresponded to molecular weight of tetramer (fractions 5-7) were negatively stained and imaged in the electron microscope (Figure 23). Image

datasets of several thousand particles were acquired for each complex (Table 6). After initial refinement attempts, the datasets were complemented by images taken at 20º tilt to improve the angular coverage. Particles were picked, cut out from micrographs, normalized and scaled to a box size of 40x40 and 4.66 Å/px.

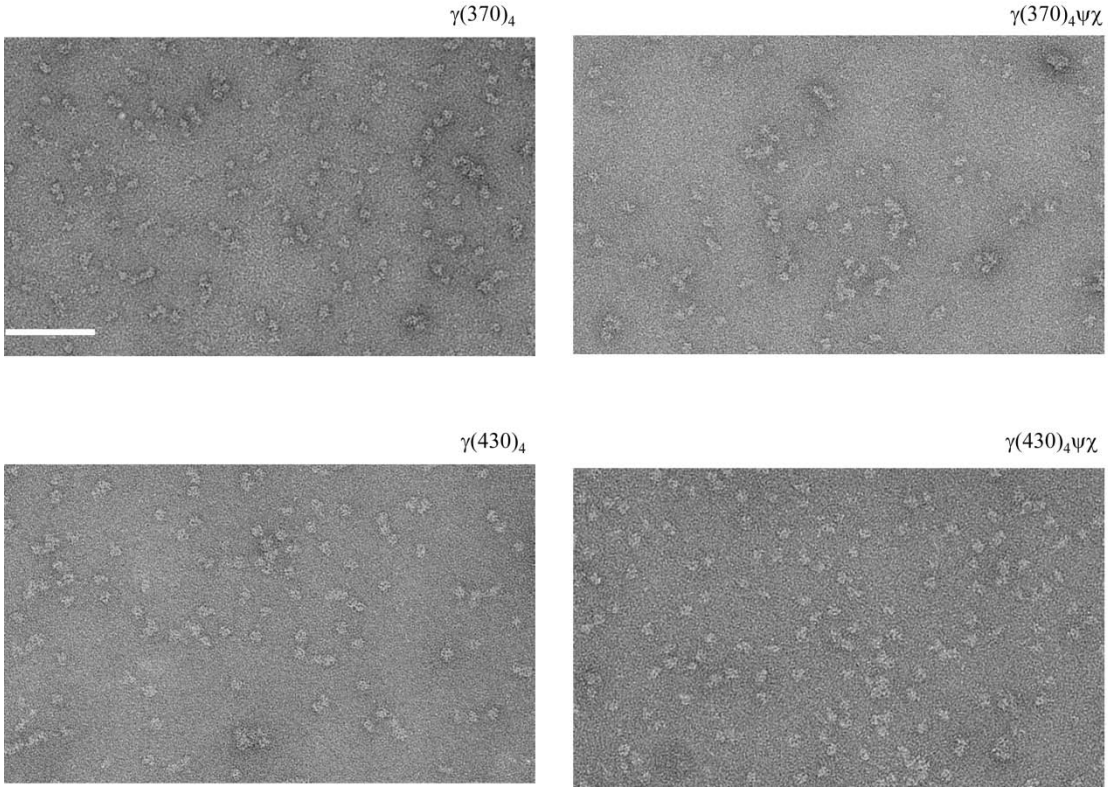


Figure 23. Micrographs of GraFix preparations of $\gamma(370)_4$, $\gamma(370)_4\psi\chi$, $\gamma(430)_4$ and $\gamma(430)_4\psi\chi$. Scale bar 100 nm.

Table 6. Number of particles selected from micrographs taken at $0^0$and 20^0angle for each complex.

	Number of selected particles		
	0 ^{0}tilt	**20 ^{0} tilt**	**Total**
$\gamma(370)_4$	10482	5623	16105
$\gamma(370)_4\psi\chi$	9217	4079	13296
$\gamma(430)_4$	13672	2890	16562
$\gamma(430)_4\psi\chi$	10378	6080	16458

Each dataset was aligned and classified in 2D using MLF2D and different high resolution cutoff values. A resolution cutoff of 35 Å provided the best defined class averages (Figure 24). Similar projection averages were found for all 4 complexes Some of the class averages showed four density blobs, consistent with the presence of tetrameric species in the complexes (labeled with asterisks in Figure 24); some were reminiscent of projections of the clamp loader crystal structure (labeled with hashes in Figure 24). All these hinted to a reasonable structural preservation; however, as before, attempts to proceed to 3D refinement failed to converge.

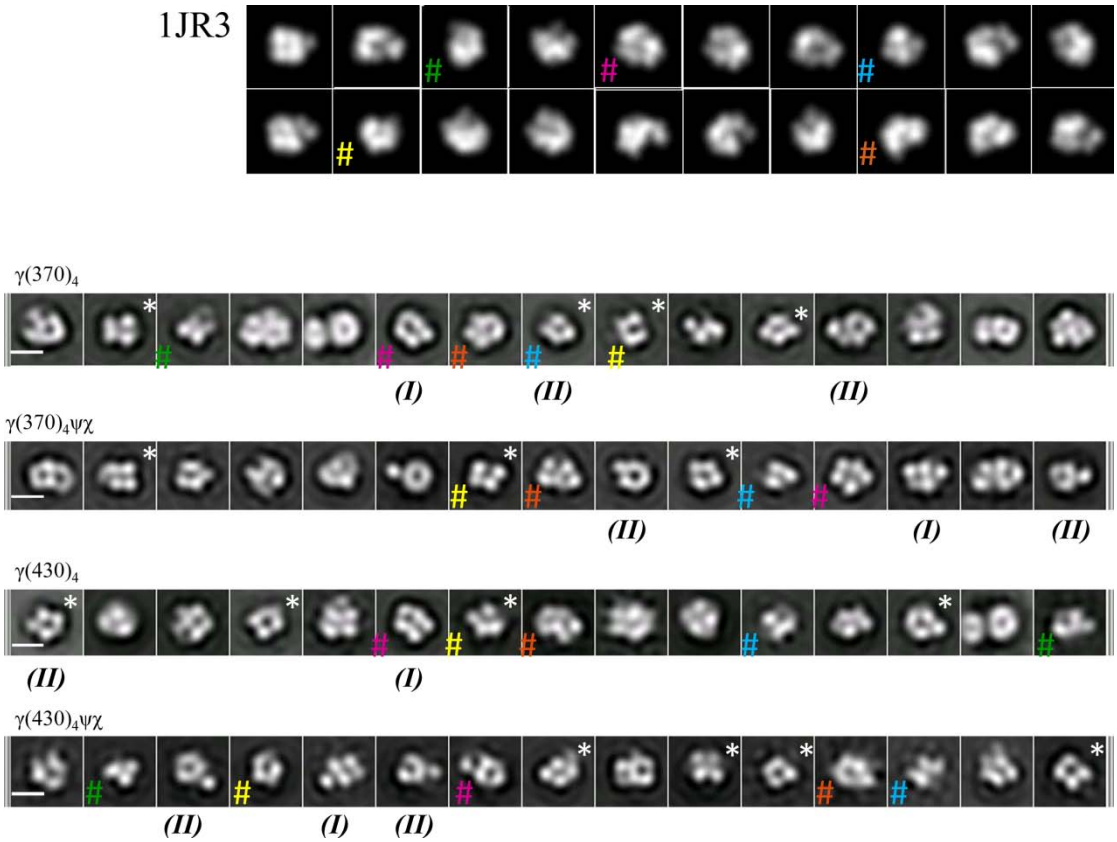


Figure 24. Class averages obtained by simultaneous 2D alignment and classification in 15 groups using MLF2D with high resolution cutoff at 35 Å. (*) tetrameric views, (#) resembling to 1JR3 projections, (I) and (II) similarities between all complexes. The scale bar represents 10nm.

4.5 Selection of homogenous particle populations ("dataset cleaning")

Since attempts to increase the homogeneity of the dataset by chemical methods (cross-linking) had failed to produce the desired results, we next investigated the possibility of carrying out this task in a computational way, using image classification

methods to isolate a self-consistent image subset. The general idea was to carry out several independent 3D refinements and classifications starting from different initial volumes, and follow the fate of the experimental projections in each of this classifications; particles grouped together when classified using different initial models should belong to a self-consistent image set.

The procedure was carried out using the data collected from GraFix preparations (Table 6). For each complex under study, three rounds of classification were performed as follows:

1st round.

Two independent runs of simultaneous 3D orientation search and classification using MLF3D (Scheres, et al., 2007) were performed. Each run started with one of two initial models: either RCT, or PDB. The initial model was low pass filtered to 80 Å resolution and its grayscale values normalized within the same range as those present in the experimental images. The complete image dataset was randomly split in three subsets, and a single iteration of MLF3D was run for each subset, using the low pass filtered initial model as a reference. This resulted in the generation of three numerically different seeds for the subsequent alignment and classification of the whole dataset in three groups. All MLF3D runs were performed including data only up to 35 Å resolution and through a maximum of 25 iterations, unless convergence was reached first. We selected the largest group of particles classified together in both runs and discarded the rest (Figure 25).

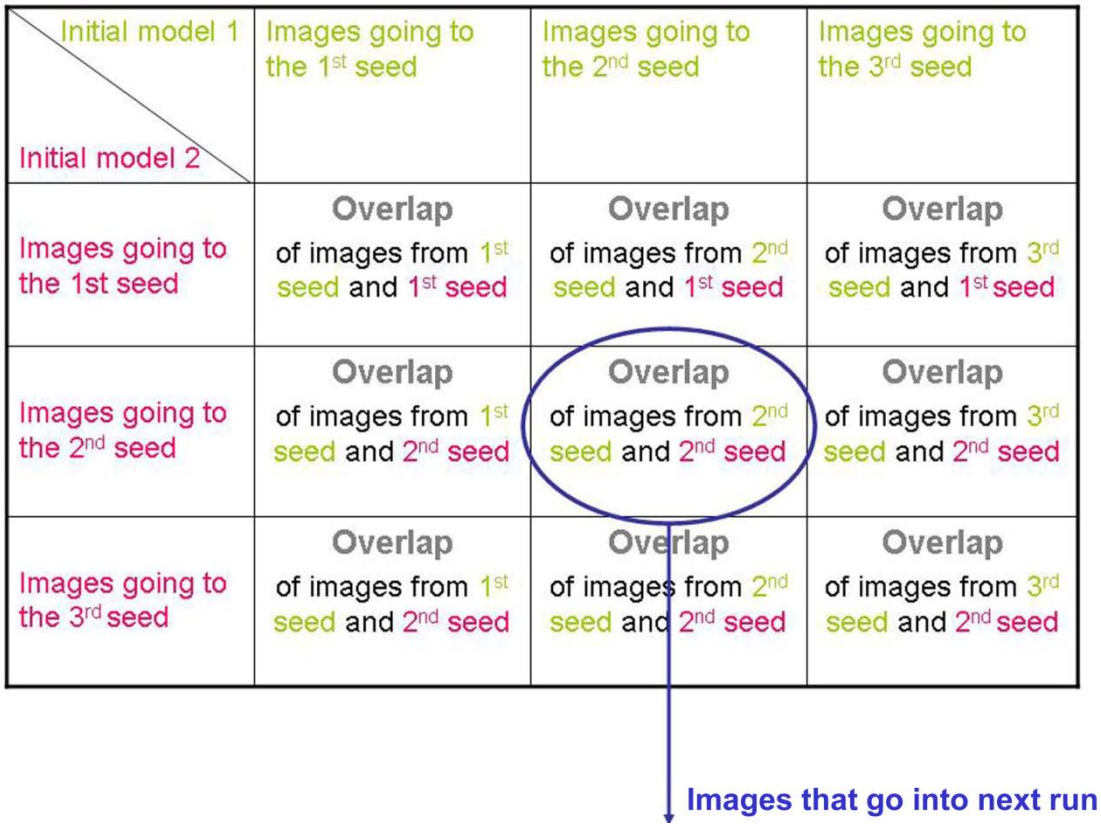


Figure 25. Determination of homogeneous datasets by selecting particles classified together in two independent MLF3D classifications with different starting models. For each tested initial model, three seed maps are created and the image dataset is simultaneously aligned and classified into three groups. Then, the overlap between classes obtained from the different classification experiments is analyzed to select the particle subset going into the next image processing step.

2nd round

The above procedure was repeated with the reduced dataset. In the second round usually there was not a majoritary group of particles clustered together, so as a criterion to further clean the dataset we discarded particles assigned to groups yielding outlier 3D maps. Figure 26 shows the process of the 1st and 2nd rounds of classification for the $\gamma(370)_4$ dataset. Figures for the equivalent process for the other complexes can be found in Appendix I.

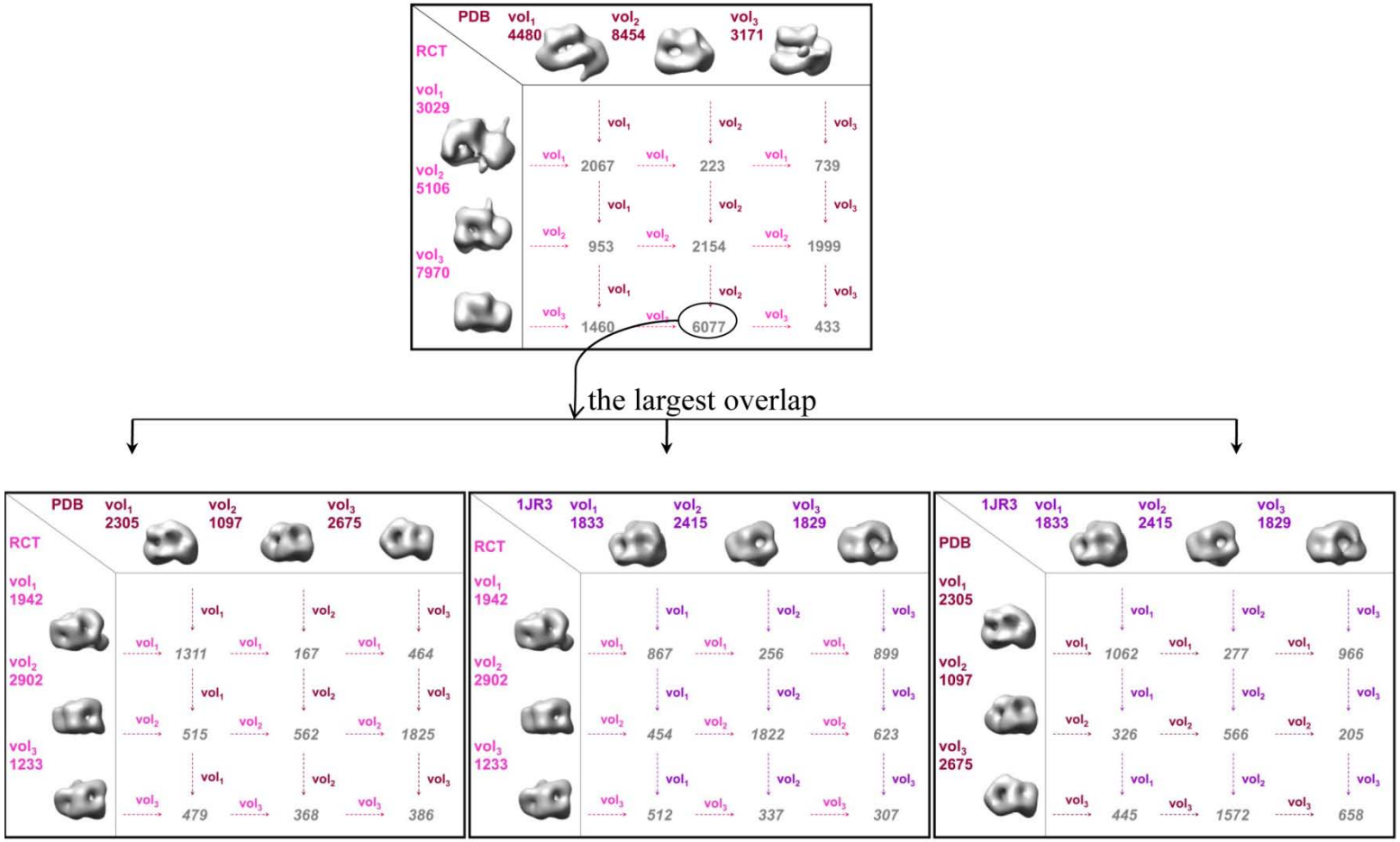


Figure 26. First (upper panel) and second (lower panels) round of MLF3D based dataset cleaning for $\gamma(370)_4$. *Violet, pink* and *dark red* figures indicate the number of images classified as most likely being projections of each particular seed. Figures *in grey* indicate the number of images classified together for each pair of seeds coming from independent MLF3D runs with different initial references. The maps shown are the final results of each MLF3D run.

3rd round

The image subset remaining after the 2nd round of classification was refined using EMAN, with two different initial models. The resulting maps (still not converging) were used as initial references to repeat the 1st and 2nd round of classification, restarting with the complete image dataset. The difference between this 3rd round and the two previous ones is that here the initial models were closer to the dataset being processed, as they had been originated from a part of this same dataset. (Figure 27, and Appendix I).

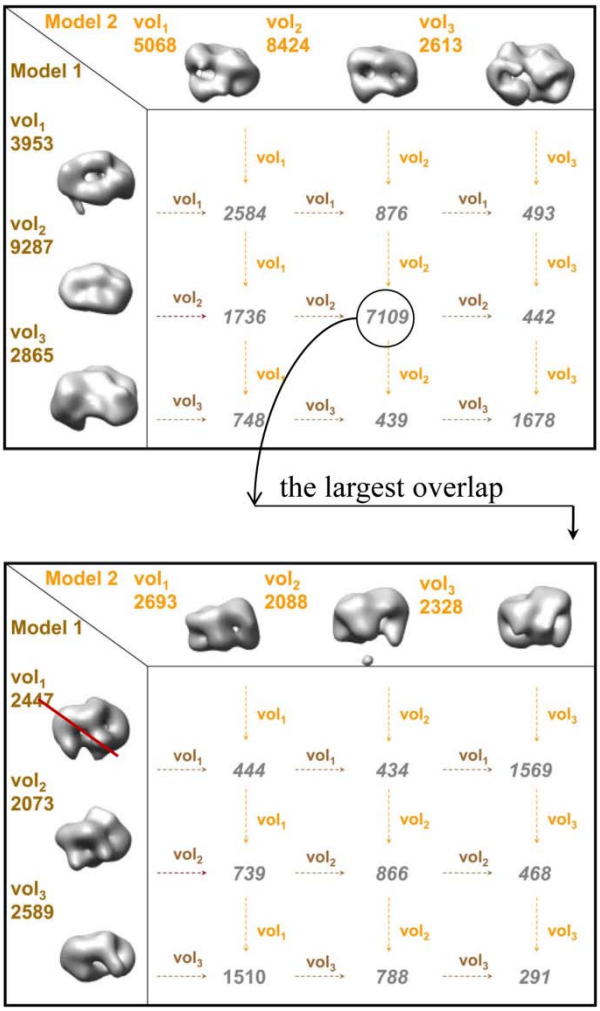


Figure 27. 3rd round of dataset cleaning for $\gamma(370)_4$. The complete dataset was used as input to repeat the previous 1st and 2nd rounds with initial maps calculated form a subset of the data. Images that went to the crossed out 3D map were not taken to further analysis.

The group of images selected after the 3rd round of dataset cleaning was considered the final subset and refined using MLF3D with only one seed against each different initial models (RCT, PDB, 1JR3, initial maps from the 3rd round). The resulting maps from these runs were aligned, and their average was used as initial reference for the final MLF3D refinement with a single seed, producing the final map for each complex. For all complexes, only 3/4ths of the initial dataset was included in the final map calculation (Figure 28).

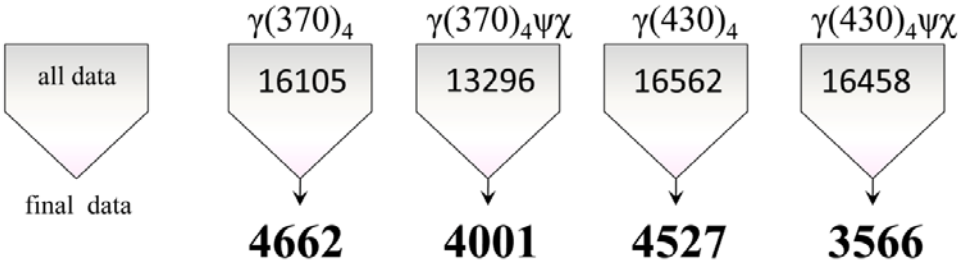


Figure 28. Initial and final number of particles remaining after dataset cleaning for all complexes.

[see comments in Appendix I]

4.6 Map validation

Several analyses were carried out to assess whether the final maps were *bona fide* representations of the object generating the experimental projections, and not just a result of model bias.

4.6.1 Cross-validation

For each complex, the final MLF3D refinement with a single seed was repeated twice: one with an initial model coming from γ_4 data, and a second one with an initial model coming from experimental $\gamma_4\psi\chi$ data. The final maps coming from both refinements should be similar. Figure 29 shows the resulting maps for all complexes. All maps were aligned, and then the pair of maps for each complex were visually and quantitatively compared. We used as a criterion for quantitative comparison the cross-correlation coefficient given by UCSF Chimera *Fit Map* tool. The cross-validation refinements produced maps similar in their general features, but the similarity was lower for some complexes, indicating a larger heterogeneity. The

most homogeneous complex was found to be $\gamma(370)_4$, while the most heterogeneous was $\gamma(430)_4$ (Figure 29, 30).

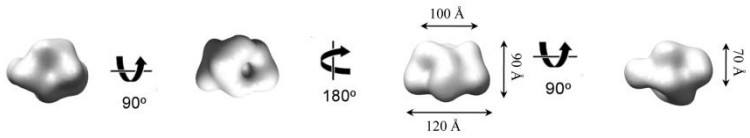
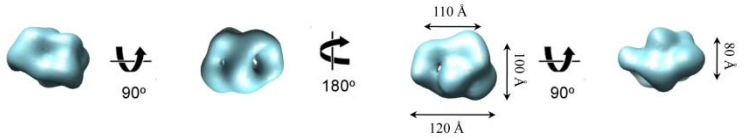
COMPLEX	(#)	Top view	Front view	Back view	Bottom view	CC
$\gamma(370)_4$	γ_4					0.95
	$\gamma_4\psi\chi$					
$\gamma(370)_4\psi\chi$	γ_4					0.71
	$\gamma_4\psi\chi$					

Figure 29. Final maps and cross validation of $\gamma(370)_4$ and $\gamma(370)_4\psi\chi$ complexes For each complex, two maps resulting from MLF3D refinement of the final image dataset against two initial models are shown from several points of view. Particle dimensions are indicated, (#) dataset of origin for initial model.

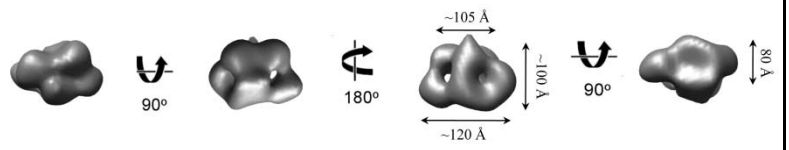
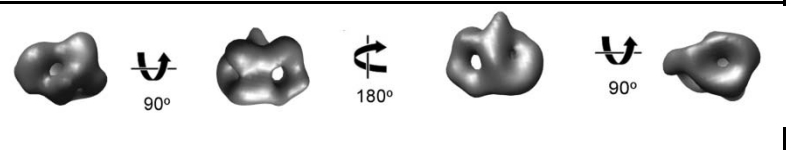
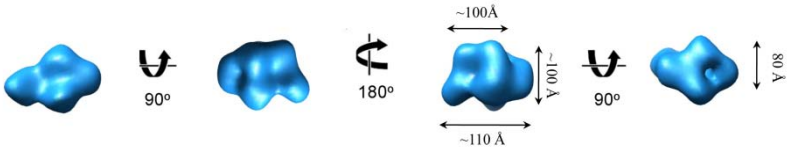
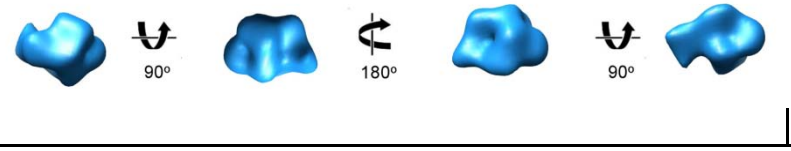
COMPLEX	(#)	Top view Front view Back view Bottom view	CC
$\gamma(430)_4$	γ_4		0.61
	$\gamma_4\Psi\chi$		
$\gamma(430)_4\psi\chi$	γ_4		0.86
	$\gamma_4\Psi\chi$		

Figure 30. Final maps and cross validation for $\gamma(430)_4$ and c $\gamma(430)_4\psi\chi$ complexes. For each complex, two maps resulting from MLF3D refinement of the final image dataset against two initial models are shown from several points of view. Particle dimensions are indicated, (#) dataset of origin for initial model.

4.6.2 Comparison with 2D class averages

As a second validation criterion, projections of the final 3D maps were compared with 2D class averages of the image dataset. 2D class averages were calculated using a reference free procedure, and as such represent the different projections present in the experimental dataset without any model bias. On the one hand, the final 3D maps were projected in all possible directions with an angular step of 10º. On the other, the final ("clean") image dataset was aligned and classified in 2D using MLF2D with 15 classes. Finally, 2D class averages were automatically aligned and compared to the projections, to find the one most similar to each 2D average. The results are presented in Figure 31. A good correspondence was found in all cases between the 3D map projections and the bias-free class averages.

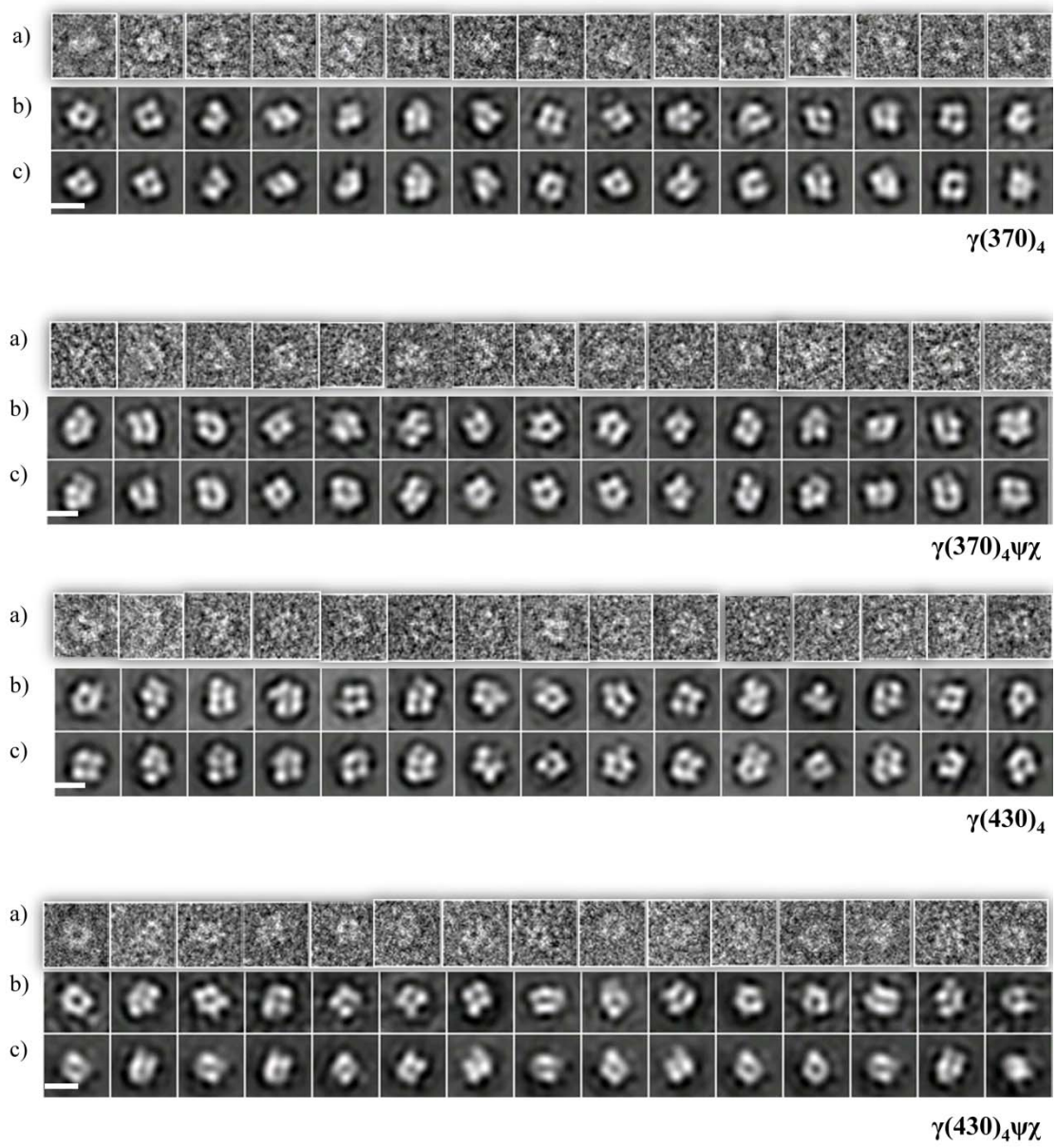


Figure 31. Comparison of final 3D map projections with 2D class averages. For each complex, row (a) shows a representative experimental image assigned to each particular class average; 2D class averages are shown in row (b), and projections of the corresponding 3D maps in row (c). Scale bar stands for 10 nm.

4.6.3 Comparison with RCT models

An additional validation criterion was to compare our final 3D maps with the collection of RCT reconstructions, also free of model bias. Although those were calculated from "unclean" datasets, it was found that some of them were similar to our final maps (Figure 32), and so this indicates that the latest represent at least some of the conformational states present in our samples.

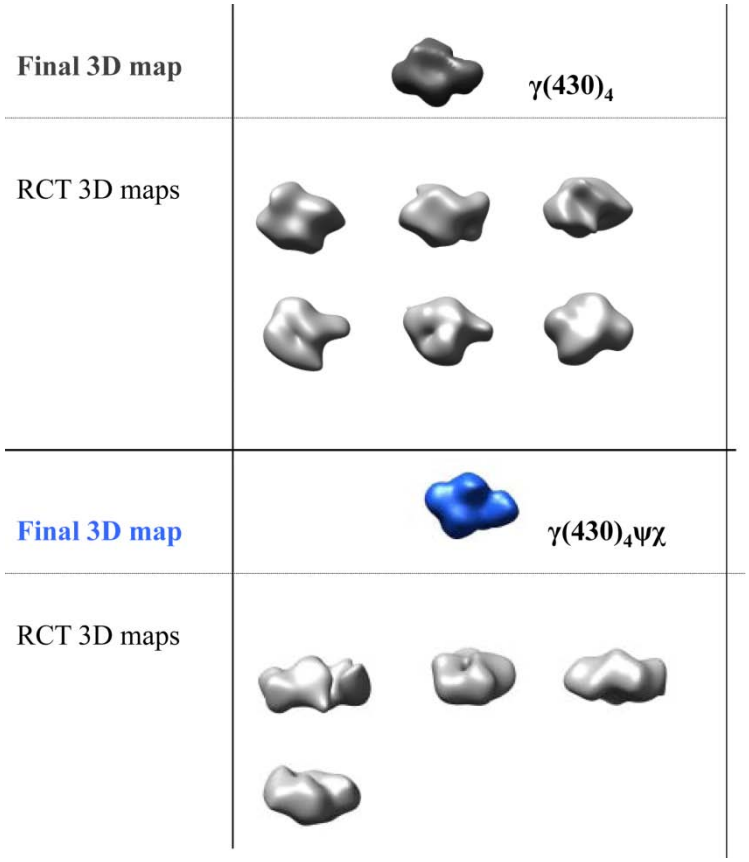


Figure 32. Comparison of RCT volumes and final 3D maps.

4.6.4 Cryo-EM analysis

To further assess our negative staining maps, we sought to obtain and analyze a cryo-EM dataset, under the hypothesis that the persistent heterogeneity found in the samples might be a negative staining artifact. GraFix fractions of $\gamma(370)_4$ complex were vitrified and imaged under cryo conditions (Figure 33) to obtain a dataset of 12311 particles. These were aligned and classified using MLF2D. Noise alignment was a problem in all our analyses, and was even worse for cryoEM data; in this case we needed to limit the resolution of data taken into account for MLF2D to 40 Å to obtain interpretable class averages. After this first analysis, particles assigned to class averages representing only background noise, or neighboring particles too close together, were discarded. The remaining 7807 particles were again classified in 2D (Figure 33b).

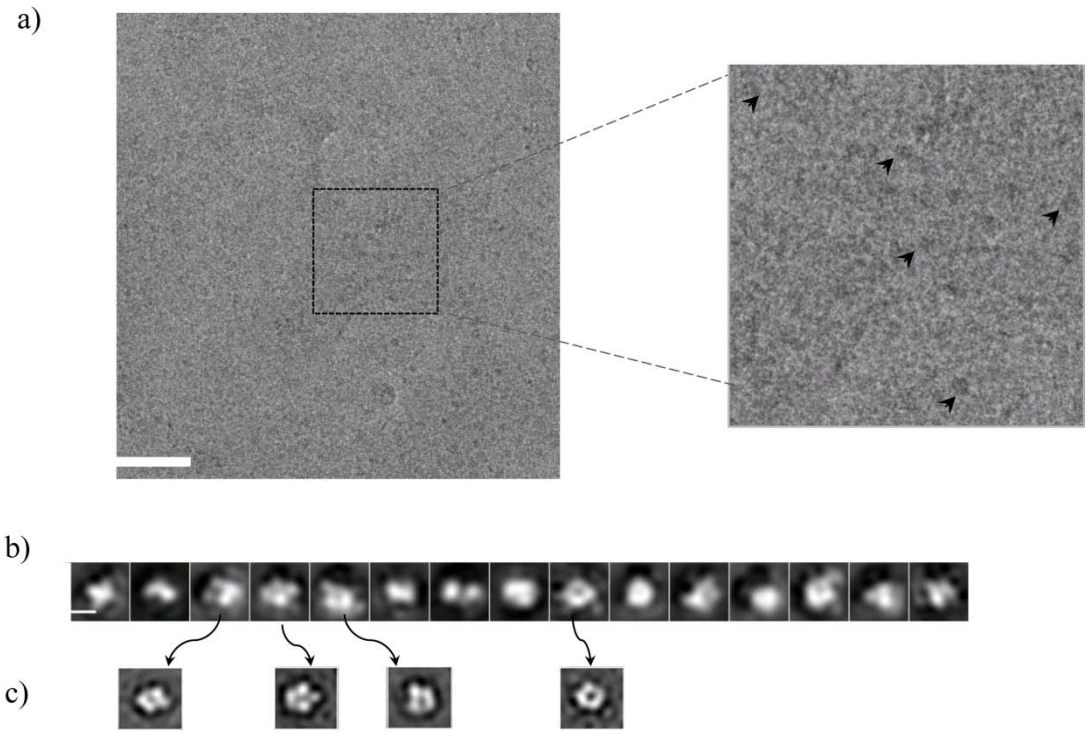


Figure 33. Cryo-EM analysis of $\gamma(370)_4$ complex. (a) representative micrograph. (b) MLF2D classification of the new dataset (c) similar classes of negative staining. Scale bars represent 100 nm in (a) 10nm in (b).

Some of the cryoEM class averages were similar to those found in negative staining preparations (Figure 33c), indicating that those represented genuine sample conformations. However, 3D refinement of the cryoEM dataset, performed as described for the negative staining images, failed to converge. This indicates that structural heterogeneity is a characteristic of the specimen and not a result of negative staining preparation.

4.7 Map interpretation

4.7.1 Map resolution

Since all maps were calculated from MLF3D refinements including data only up to 35 Å in the probability calculations, this is the range of resolution attained. We estimate the resolution of each map as the highest frequency for which MLF3D gave a non-zero spectral signal to noise ratio (SSNR) value. The values thus obtained were: 31 Å for $\gamma(370)_4$, $\gamma(370)_4\psi\chi$ and $\gamma(430)_4$; and 37 Å for $\gamma(430)_4\psi\chi$. Attempts to refine

the maps including higher frequencies resulted in artifacts due to background noise alignment.

4.7.2 Difference maps

Difference densities were calculated by subtracting maps of different complexes that had been refined against a common initial reference, aligned and normalized, as described in Section 3.9. Subtraction of $\gamma(370)_4$ from $\gamma(370)_4\psi\chi$ produced a strong difference peak with a size consistent with that of the $\psi\chi$ heterodimer (Figure 34a). However, no such peak was observed in the $\gamma(430)_4\psi\chi$ minus $\gamma(430)_4$ difference map (Figure 34b). Similarly, no clear, interpretable peak was obtained for the 60 C-terminal residues of γ in the $\gamma(430)_4$ minus $\gamma(370)_4$ difference (Figure 34c).

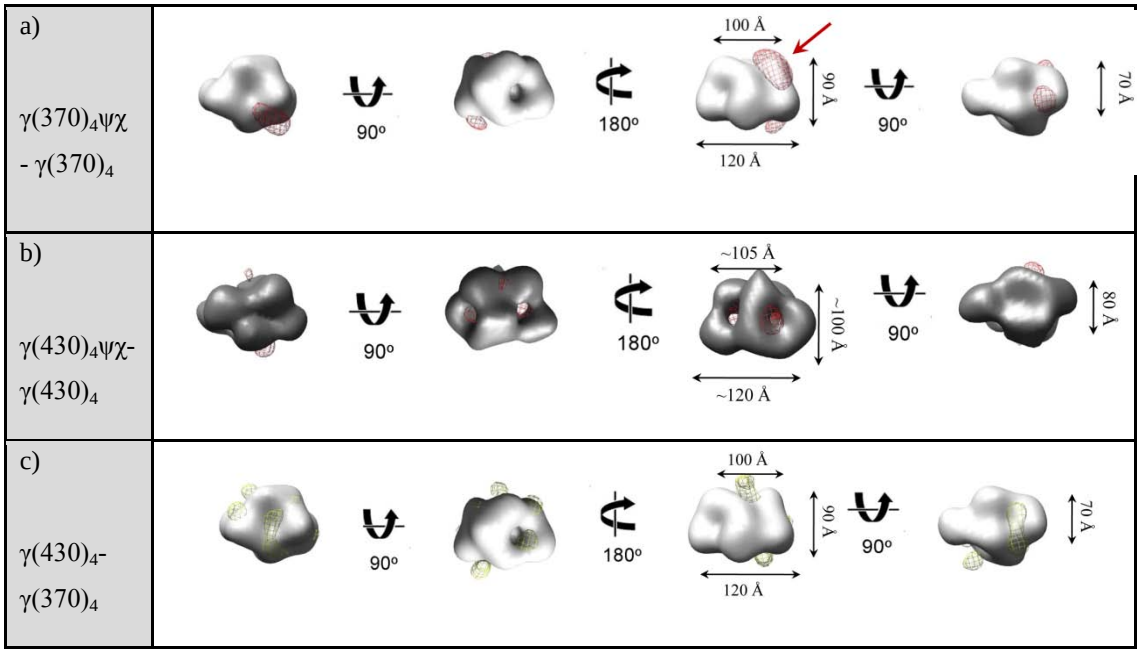


Figure 34. Difference maps calculated to locate: (a) $\psi\chi$ in the $\gamma(370)_4\psi\chi$ complex; (b) $\psi\chi$ in the $\gamma(430)_4\psi\chi$ complex; and (c) C-terminal region of γ The $\gamma(370)_4$ map is shown in light gray; $\gamma(430)_4$ in dark gray. The difference densities are shown in yellow .Complex maps and difference maps are rendered at a threshold of 3σ standard deviations. A strong difference peak consistent with the size of $\psi\chi$ is indicated with an arrow in (a).

4.7.3 Fitting of crystal structures

The crystal structures of four identical γ(370) monomers (taken from the clamp loader structure, PDB ID 1JR3) and of the ψχ heterodimer (PDB ID 1EM8) were fitted to our EM maps to obtain additional information on the architecture of the complexes. The central γ monomer in the clamp loader, which contacts only other γ monomers, was chosen for the fitting. Although in the clamp loader the three γ molecules adopt different conformations, they are highly similar when filtered at the EM map resolution. Atomic structures were first fitted manually, and then their positions were refined in UROX (Siebert and Navaza, 2009) considering each γ monomer, and the ψχ heterodimer, as independent rigid bodies. To decide how the molecules should be oriented in the first manual fit, we made use of two pieces of information.

First, visual comparison revealed some orientations in which our $\gamma(370)_4$ map was very similar to the crystal structure of the clamp loader filtered at the same resolution (Figure 35). In both cases the complex side views have a trapezoidal profile, with the narrow base corresponding to the γ C-terminal collar in the clamp loader structure. We therefore hypothesized that in the γ homotetramer the C-terminal domains might also be forming the narrow base of the trapezoid.

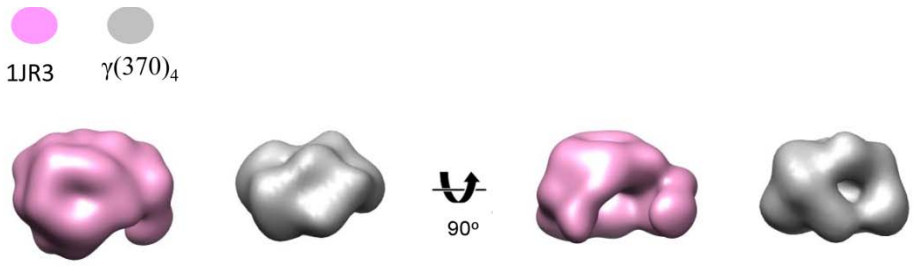


Figure 35. Representation of 1JR3 crystal structure filtered to 35 Å resolution (pink). In grey our $\gamma(370)_4$.

Second, in the $\gamma(370)_4\psi\chi$ minus $\gamma(370)_4$ difference map the peak consistent with the size of ψχ appears closer to the narrow base of the trapezoid than to the wide base. Since it has been reported that the ψχ interaction with γ takes place through the

26 N-terminal residues of ψ and the C-terminal domain of γ (Gulbis, et al., 2004; Simonetta, et al., 2009), this difference peak was consistent with our hypothesis that the narrow base of the trapeze corresponds to the C-terminal domains of γ. The ψχ dimer was therefore located in the area corresponding to the difference peak and oriented with ψ closer to the narrow base of the complex.

Some bad clashes between molecules were observed after optimization of the and $\gamma(370)_4$ fitting with UROX, and therefore the optimization was repeated allowing UROX to simultaneously refine the map scale. The best cross-correlation coefficient values were obtained when the maps were scaled by a factor of 1.07, and the clashes were relieved. Both the final cross-correlation values (0.89 for $\gamma(370)_4$ and 0.91 for $\gamma(370)_4\psi\chi$) and visual inspection indicated a good fit between the crystal structures and the EM maps (Figure 36).

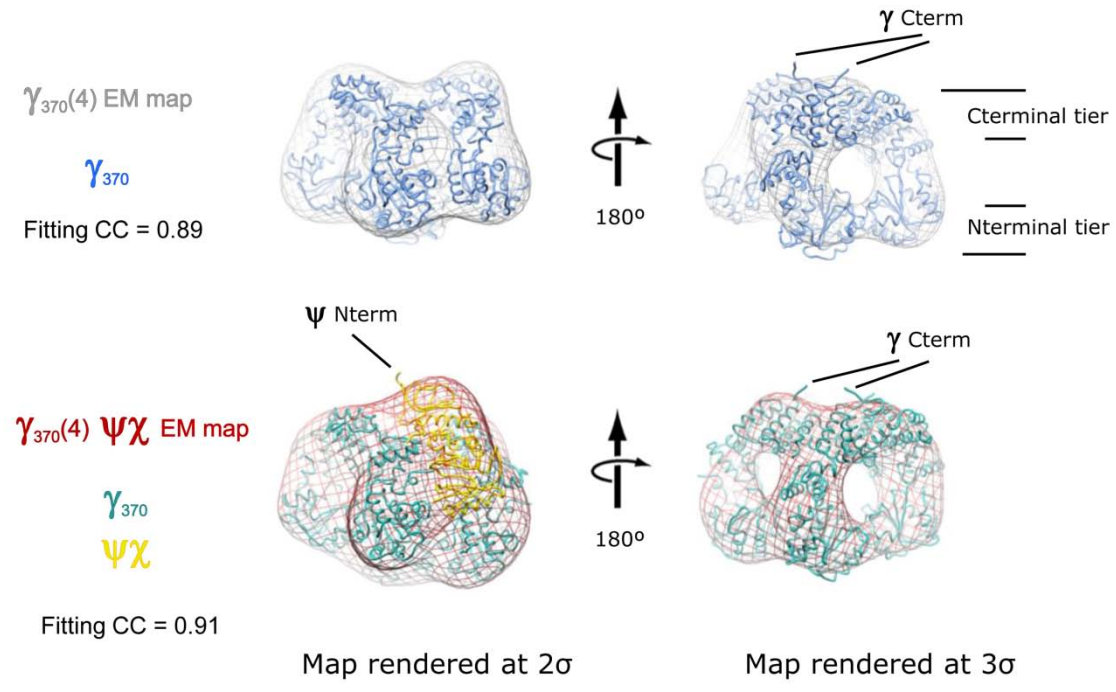


Figure 36. Two different views of crystal structures fitted to the $\gamma(370)_4$ and $\gamma(370)_4\psi\chi$ EM maps, to show quality of fit. The EM maps are shown as a semitransparent mesh, while the C-alpha trace of the fitted molecules is shown as a tube. Maps are rendered at a 2σ threshold in the left hand side, and at a 3σ threshold in the right hand side. Color key is indicated in the figure, as well as the position of some critical regions in the structures.

4.7.4 Protein-protein interactions in the γ_4 and $\gamma_4\psi\chi$ complexes

Analysis of the fitted molecules revealed the protein-protein interactions in the γ(370) homotetramer and in the $\gamma(370)_4\psi\chi$ complex. These interactions are best observed when the atomic structures are represented as low resolution surfaces (Figure 37b). To describe the organization of the γ homotetramer, we consider each monomer as a C-shaped object. Four such objects are located around the lateral facets of a distorted cube, forming an asymmetric closed circular structure (Figure 37a). The cube is organized in two tiers: the γ C-terminal domains form the upper tier, while the N-terminal domains make up the lower tier. All γ-γ interactions in the cube are unique: head-to-head, tail-to-tail, and two different kinds of head-to-tail interactions are observed (Figure 37b). This is in contrast with the organization of γ in the clamp loader, where the ring is closed by the accessory subunits δ and δ'(Figure 37 bottom-right). In the clamp loader, the γ-γ interactions vary due to the different conformations of the molecule, and some of them are different from those found in our analysis . In the terminology used here, in clamp loader they would all fall within the "side-to-side" category, that would not be possible to obtain a closed circular structure (our analyzed complexes) using only this kind of interaction..

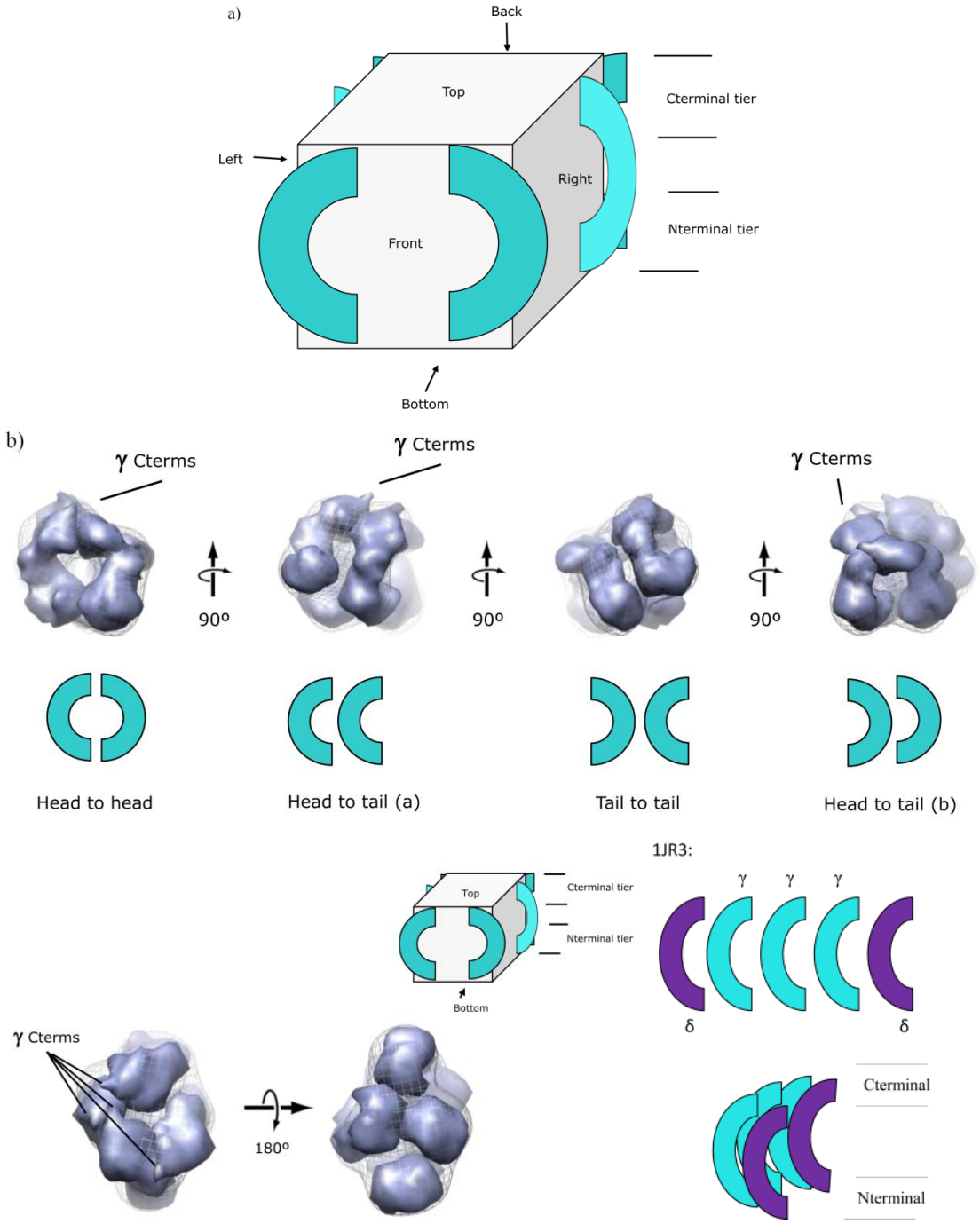


Figure 37. Interactions between monomers in the $\gamma(370)_4$ complex.(a) Schematic representation. (b) The crystal structures of four γ monomers fitted to the EM map (gray) are represented as low resolution surfaces (blue) to show the different kinds of contactsContacts in 1JR3, minimum clamp loader are shown at the bottom right side.

Fitting of crystal structures in the $\gamma(370)_4$ showed the particular position occupied by the heterodimer, and the modification in γ-γ contacts induced by its presence. ψχ interacts with the γ tetramer at a well defined location, disrupting the tail-to-tail contact (Figure 38) . This disruption makes one of the γ monomers swing inwards, thus altering one of head-to-tail contacts too.

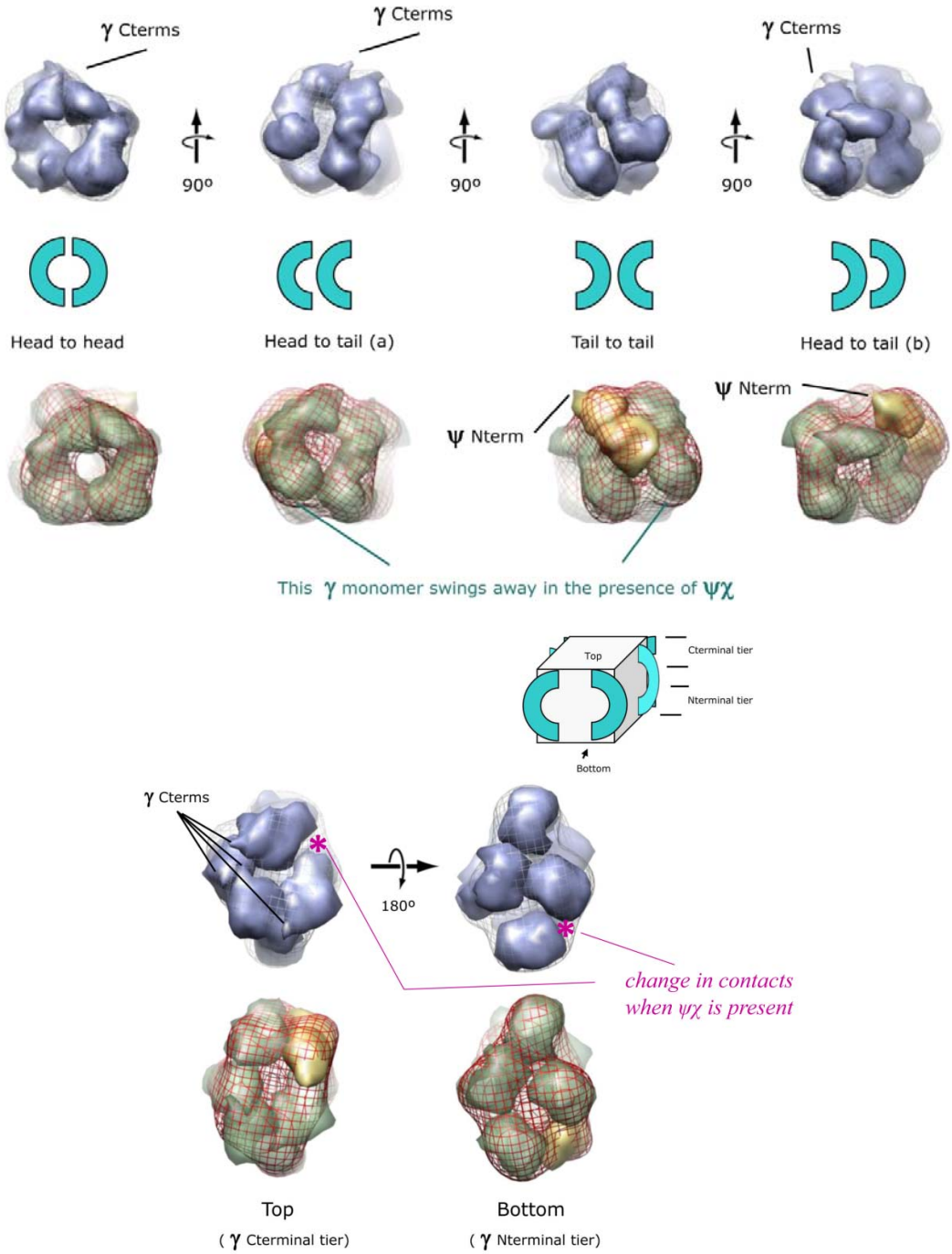
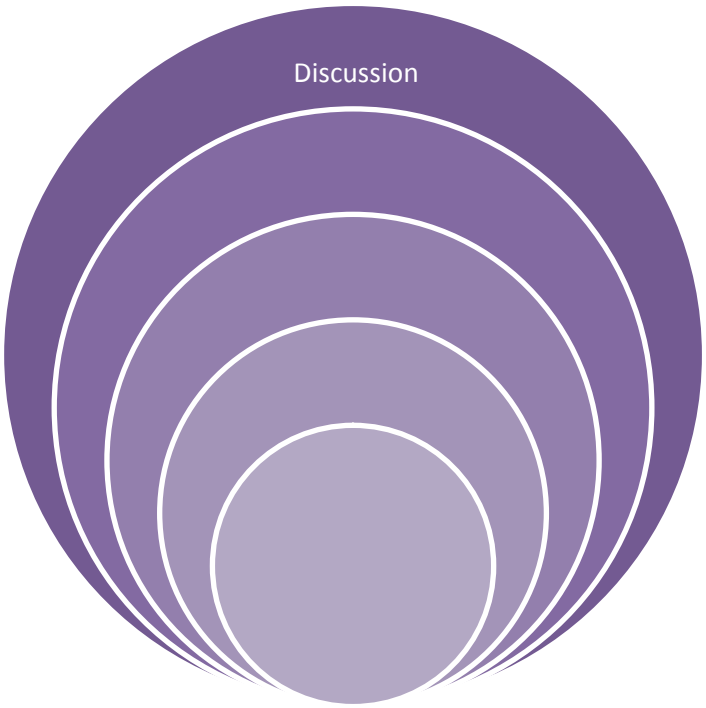


Figure 38 Location of ψχ in the $\gamma(370)_4\psi\chi$ complex. The upper and middle rows show the γ-γ contacts as in Figure 36. In the bottom row the $\gamma(370)_4\psi\chi$ EM map is shown as a red mesh, with the γ monomers in green and the ψχ dimer in yellow, in the same orientations as those used for $\gamma(370)_4$ in the top row.


Discussion

5. *Discussion*

5.1 3DEM and complex heterogeneity

In this Thesis we have used 3DEM to analyze the structure of two complexes involved in assembly of the *E. coli* DNA Pol III clamp loader: γ_4, and $\gamma_4\psi\chi$. Two variants of each complex have been analyzed: one containing the 430 residue (full length γ), the other containing a 370 residue long (C-terminal truncated version of γ). Although the preparations were pure and homogeneous (in the sense of giving a single peak in gel filtration experiments), all displayed large structural heterogeneity when analyzed by EM. Sample cross-linking reduced the heterogeneity, but it did not completely remove it, so that 3D refinements could not reach convergence. To finally reach a reasonable convergence, we had to resort to stringent dataset cleaning (discarding about 75% of images in each dataset) and image processing using only low frequency data (up to 35 Å).

Even after identifying a reduced, self-consistent dataset and obtaining an experimental 3D map, it was not possible to pursue the refinement to even moderately higher resolution (25 Å). This may be due to some mixture of conformations still present in our final dataset. In fact, an inverse correlation was found between complexity of the specimen and goodness of 3DEM convergence in cross-validation experiments: the best convergence was reached by the truncated $\gamma(370)_4$ complex, lacking the C-terminal region, while both the presence of full length γ or of ψχ deteriorated convergence The C-terminal region of γ is predicted to be unstructured when analyzed with FoldIndex (Prilusky, et al., 2005) or Modbase (Pieper, Eswar, Webb, Eramian, Kelly, Barkan, Carter, Mankoo, Karchin, Marti-Renom, Davis and Sali, 2009). Previous reports indicate that while γ forms stable tetramers, binding of ψχ moderately shifts the equilibrium inducing the apparition of γ trimers (Park, et al., 2010). Both flexibility of γ C-terminal regions and tetramer disruption by ψχ can cause further conformational variability in the complexes.

An additional factor resulting in lack of convergence could be the intrinsic difficulty in aligning the experimental projections produced by this kind of complexes. Our specimens are in the smallest size range suitable for EM analyses (<200 kDa) (Henderson, 2004), therefore the signal to noise ratio in their projections is extremely low, even after negative staining. They are globular, almost cubic, but without symmetry or conspicuous features to help in the orientation search. These two factors resulted in artifacts due to background noise alignment in all attempts to increase the resolution in our image processing.

In spite of the mentioned difficulties, we believe that the final maps do represent the structure of the studied complexes – at least of a part of the different conformations that the complexes can adopt. This belief is supported by the good match between projections of the final maps and reference-free 2D class averages of the experimental data; by cross-validation with refinements coming from different initial models; and by the finding of RCT maps (also reference free) similar to the final maps obtained.

Structural variability may have arisen from intrinsic complex flexibility, polydispersity, or from distortions caused by the negative staining sample preparation process. The presence of high levels of heterogeneity in cryoEM data analysis indicates that this is not the case. We cannot rule out the possibility that interaction with the carbon support is producing the structural distortions, instead of the staining agent. Because of the small size of our complexes, the cryoEM preparation used a thin carbon layer, since without it it is extremely difficult to obtain an ice layer so thin to actually have enough contrast to see the particles. New EM techniques not relying on phase contrast, such as cryoEM with Zernike plates, have been proposed for improving data acquisition for these small complexes (Shigematsu, Sokabe, Danev, Tominaga and Nagayama, 2010). However this technique is still not widely available.

Other structural biology techniques might help to shed light on the structure of γ_4 and $\gamma_4\psi\chi$. Protein crystallography has been used to solve the minimal clamp loader structure $\gamma_3\delta\delta'$ in a variety of conditions (Jeruzalmi, et al., 2001; Podobnik, Weitze, O'Donnell and Kuriyan, 2003; Simonetta, et al., 2009). Although the structural

variability we observe indicates that crystallization would not be a trivial task for γ_4 or $\gamma_4\psi\chi$, at least one of them has already been crystallized (X. S. Chen & C. S. McHenry, personal communication). However, attempts to phase the diffraction data using molecular replacement or our EM maps have been unsuccessful so far. It is possible that the particular conformation(s) of γ forming the crystals is different from either the previously solved crystal structures or our oligomeric complexes. Another alternative is the use of small angle X-ray diffraction (SAXS). A fundamental strength of SAXS is that it provides the overall structure including both architectural arrangements and conformations in the 50–10 Å resolution range in near physiological conditions.

Solution scattering is continuous and radially symmetric (isotropic), in contrast to X-ray crystallography where the molecules are highly organized within a crystal lattice (Putnam, Hammel, Hura and Tainer, 2007). SAXS has been successful in providing structural information on specimens refractory to other structural biology techniques, such as p53(Tidow, Melero, Mylonas, Freund, Grossmann, Carazo, Svergun, Valle and Fersht, 2007) and αβ-crystallin (Jehle, Rajagopal, Bardiaux, Markovic, Kuhne, Stout, Higman, Klevit, van Rossum and Oschkinat, 2010).

5.2 Structure of the complexes

All complexes analyzed form a closed circular structure, but they present no cyclic symmetries, not even in the homotetramers. This lack of symmetry in homo-oligomeric structures has previously been observed for other ring loading complexes, such as *B. subtilis* DnaB (Nuñez-Ramirez, Velten, Rivas, Polard, Carazo and Donate, 2007). Moreover, our results indicate that γ monomers can establish many different sorts of interactions between them, and some of the interactions observed in the homotetramer are the same established by γ in the clamp loader, when δ and δ' are present. In the clamp loader, the γ-γ interactions are tighter and position the three γ copies in such a way that addition of a fourth one would not result in a closed ring.

It has recently been reported that δ', and only δ', is responsible for disrupting the γ/τ tetramer to generate the trimer that willl then bind δ to form the clamp loader

(Park, et al., 2010). It has been proposed that this oligomer breaker activity of δ' is related to its structural rigidity (Johnson and O'Donnell, 2005). Binding of δ' would lock γ in the conformation needed to assemble into the clamp loader. This hypothesis correlates well with our results. We cannot model different γ conformations due to the limited resolution of our maps, but we can observe how γ has an extremely large versatility to establish protein-protein interactions, even with other identical molecules. This versatility is in agreement with replisome proteins, and the polymerase in particular, being intrinsically dynamic to accomplish the different stages of DNA replication. Binding of δ' may disrupt one of particular γ-γ interface, sequestering some surface contacts and exposing others that would guide assembly towards the side-to-side interactions seen in the clamp loader. Similarly, the first γ-γ binding event, which might occur in a variety of ways, might select for of certain type of interaction and leave the others free to complete the tetramer, resulting in a variable, asymmetric structure even though formed by four identical molecules.

It has also been observed that ψχ has a limited oligomer breaker activity, producing a small amount of γ_3ψχ complexes when incubated with the γ tetramer (Park, et al., 2010). However this does not prepare γ to bind δ and assemble the clamp loader. Again, this can be correlated to our results. We do not see drastic changes in the overall structure of the γ tetramer when bound to ψχ, but our model obtained by fitting atomic coordinates indicates that ψχ disrupts the tail-to-tail interaction between γ monomers, and also one of the head-to-tail ones. This may be enough to render the tetramer unstable and trigger release of one of the γ monomers in a small proportion of complexes.

5.3 Mapping complex components: future directions

We have used difference mapping between 3DEM maps to obtain information on the organization of the different components within the analyzed complexes. We expected the difference map between $\gamma(430)_4$ and $\gamma(370)_4$ to reveal the position of the γ C-terminal domain; however, this did not happen. As discussed above, this region of γ is expected to be largely unstructured; as such, it may adopt a different organization in each individual complex and be averaged out during 3DEM refinements.

We did obtain a reasonable difference map between $\gamma(370)_4\psi\chi$ and $\gamma(370)_4$, accounting for the presence of $\psi\chi$. This, together with some similarity between our 3D maps and a low resolution map calculated from the clamp loader crystal structure, has allowed us to propose a model of the general architecture of the complexes. In the case of the γ full length complexes, however, no interpretable difference was observed that could be assigned to the $\psi\chi$ heterodimer. This suggests that the presence of the γ C-terminal regions produces a more delocalized, or unstable, binding of $\psi\chi$, which precludes its observation in averaged maps. It could also be that our dataset "cleaning" procedure has resulted in selecting particles lacking the $\psi\chi$ dimer in the sample.

How can our current model for the γ_4 and $\gamma_4\psi\chi$ complexes be assessed? The best proof would be specific labeling of either $\psi\chi$ or a particular domain of γ. Specific labeling for EM is usually accomplished by using antibodies. However, for such a small and structurally heterogeneous specimen, antibody labeling may result in further complicating the image processing task. A more suitable methodology for this small particles would be direct binding of small (1 nm) gold probes to particular reactive sites in the complex (Figure 39). This is a strategy that we have started to pursue, but has not reached success yet. Based on the known crystal structures, we designed a C-terminal His-tagged ψ mutant that would allow labeling with Ni-NTA Nanogold. However the His tagged protein was insoluble, both when expressed alone or together with χ (work by McHenry group).

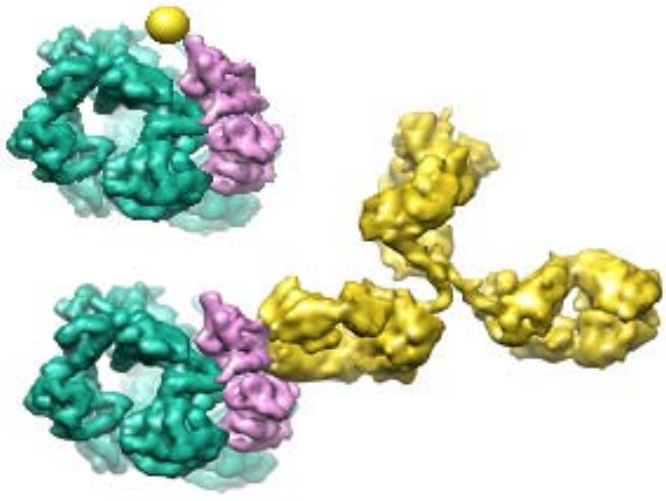
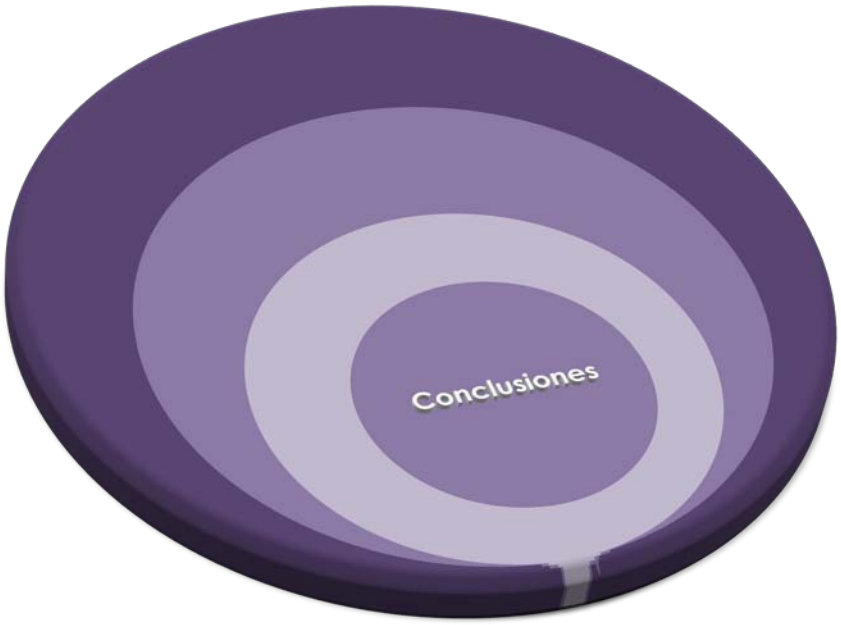


Figure 39. Visualizing of complex with a $\psi\chi$ and its gold Ni-NTA(top) and antibody (bottom) labeling results.

We also investigated the possibility of using His-tagged versions of γ to locate specific γ regions with the same probe. Since a His-tagged *E. coli* γ was not available, we tried this approach with a truncated, N-terminal His-tagged *B.subtilis* τ, the γ equivalent in the Gram-positive replisome (Martinez-Jimenez, Mesa and Alonso, 2002) (protein provided by Dr. S. Ayora). Unfortunately, Ni-NTA-Nanogold labeling assays were dominated by large protein aggregation, possibly due to the presence of multiple tags in each complex (Buchel, Morris, Orlova and Barber, 2001). Current work (underway at the McHenry laboratory) aims at obtaining complexes with His-tagged χ that may not suffer from insolubility and allow localization by Ni-NTA Nanogold.


Conclusiones

6. *Conclusions*

1. The quaternary structure of *E. coli* clamp loader subcomplexes γ_4 and $\gamma_4\psi\chi$ has been analyzed using electron microscopy and image processing.

2. A combination of chemical stabilization (cross-linking) and stringent image selection (dataset cleaning) was critical to reach consensus structures on these challenging specimens, which are small, asymmetrical, and present a high structural variability.

3. Difference maps allowed the localization of the $\psi\chi$ dimer in the truncated $\gamma(370)_4\psi\chi$, but not in the γ full length complex. Likewise, difference maps did not reveal the organization of the γ C-terminal regions.

4. Combination of crystal structures and EM maps resulted in a structural model for the $\gamma_4\psi\chi$ and γ_4 complexes.

5. The model indicates that γ has an extreme versatility to establish protein-protein interactions, even among identical monomers. This versatility is likely to be relevant for its functional activity in replisome dynamics.

6. The model also shows how $\psi\chi$ interacts with the γ tetramer at a defined position, altering certain interactions between γ monomers. This could explain the limited oligomer-breaker activity previously shown for $\psi\chi$.

7. *Appendix I*

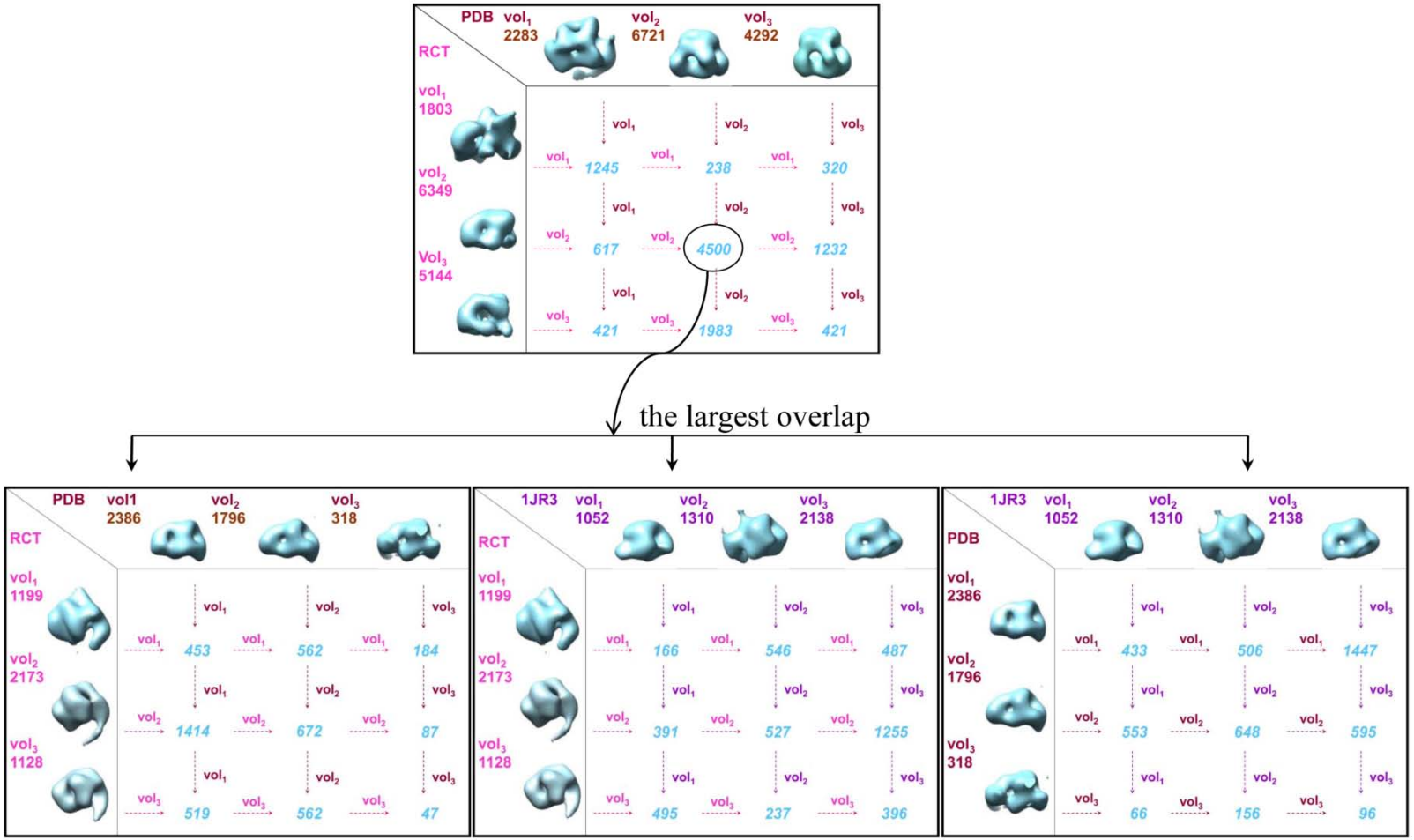


Figure 40. First (upper panel) and second (lower panels) round of MLF3D based dataset cleaning for γ(370)4 ψχ. Violet, pink and dark red figures indicate the number of images classified as most likely being projections of each particular seed. Figures in blue indicate the number of images classified together for each pair of seeds coming from independent MLF3D runs with different initial references. The maps shown are the final results of each MLF3D run..

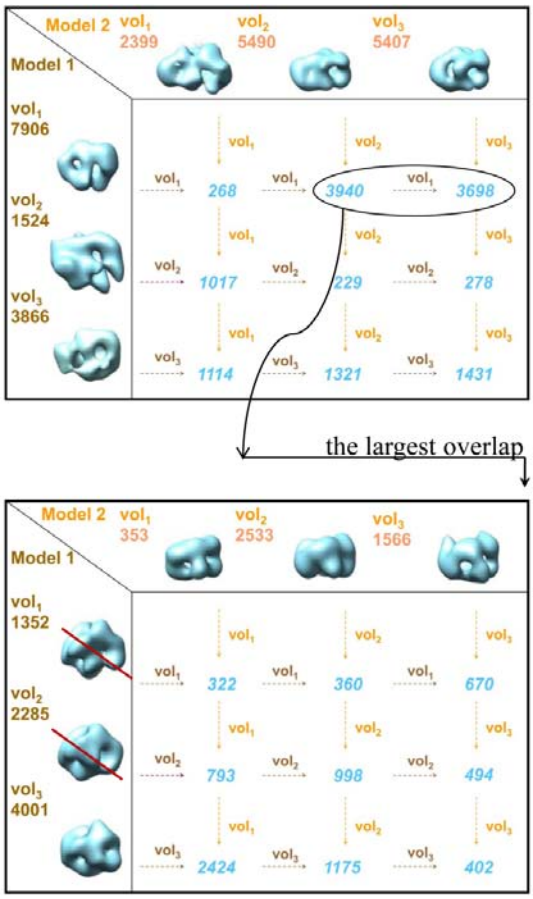


Figure 41. 3rd round of dataset cleaning for γ(370)4ψχ. The complete dataset was used as input to repeat the previous 1st and 2nd rounds with initial maps calculated form a subset of the data. Images that went to the crossed out 3D map were not taken to further analysis.

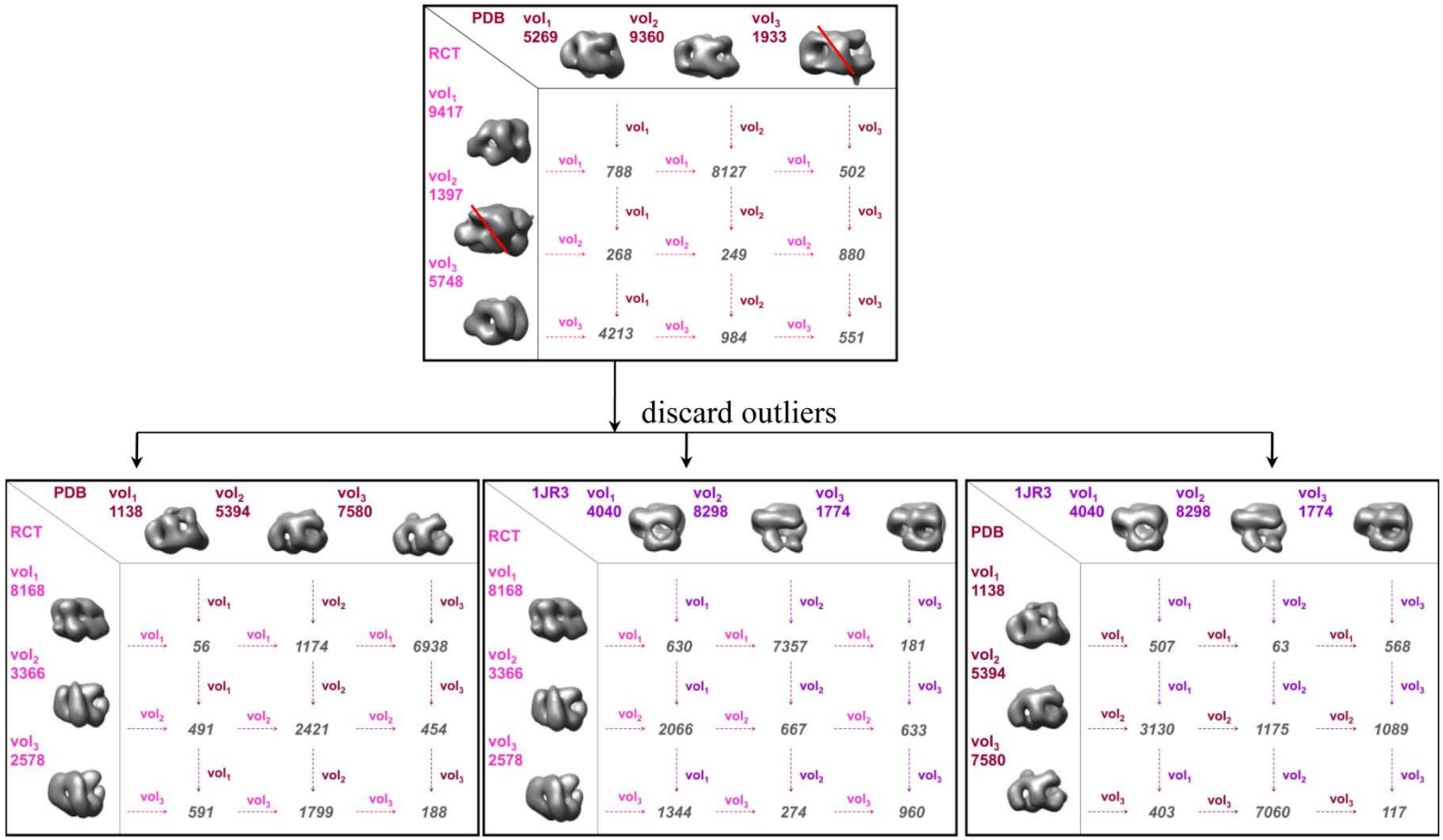


Figure 42. First (upper panel) and second (lower panels) round of MLF3D based dataset cleaning for $\gamma(430)_4$. *Violet*, *pink* and *dark red* figures indicate the number of images classified as most likely being projections of each particular seed. Figures *in dark grey* indicate the number of images classified together for each pair of seeds coming from independent MLF3D runs with different initial references. The maps shown are the final results of each MLF3D run.

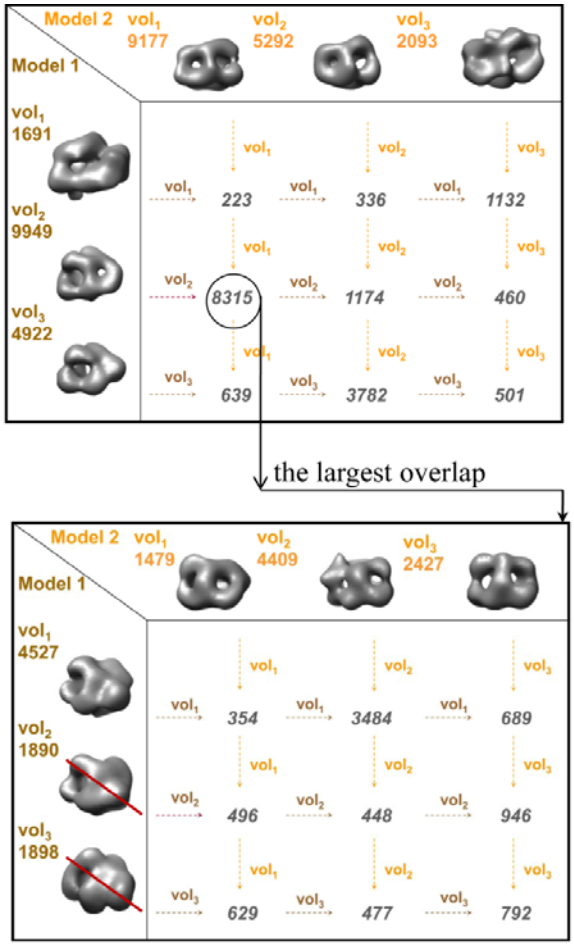


Figure 43. 3rd round of dataset cleaning for γ(430)4. The complete dataset was used as input to repeat the previous 1st and 2nd rounds with initial maps calculated form a subset of the data. Images that went to the crossed out 3D map were not taken to further analysis.

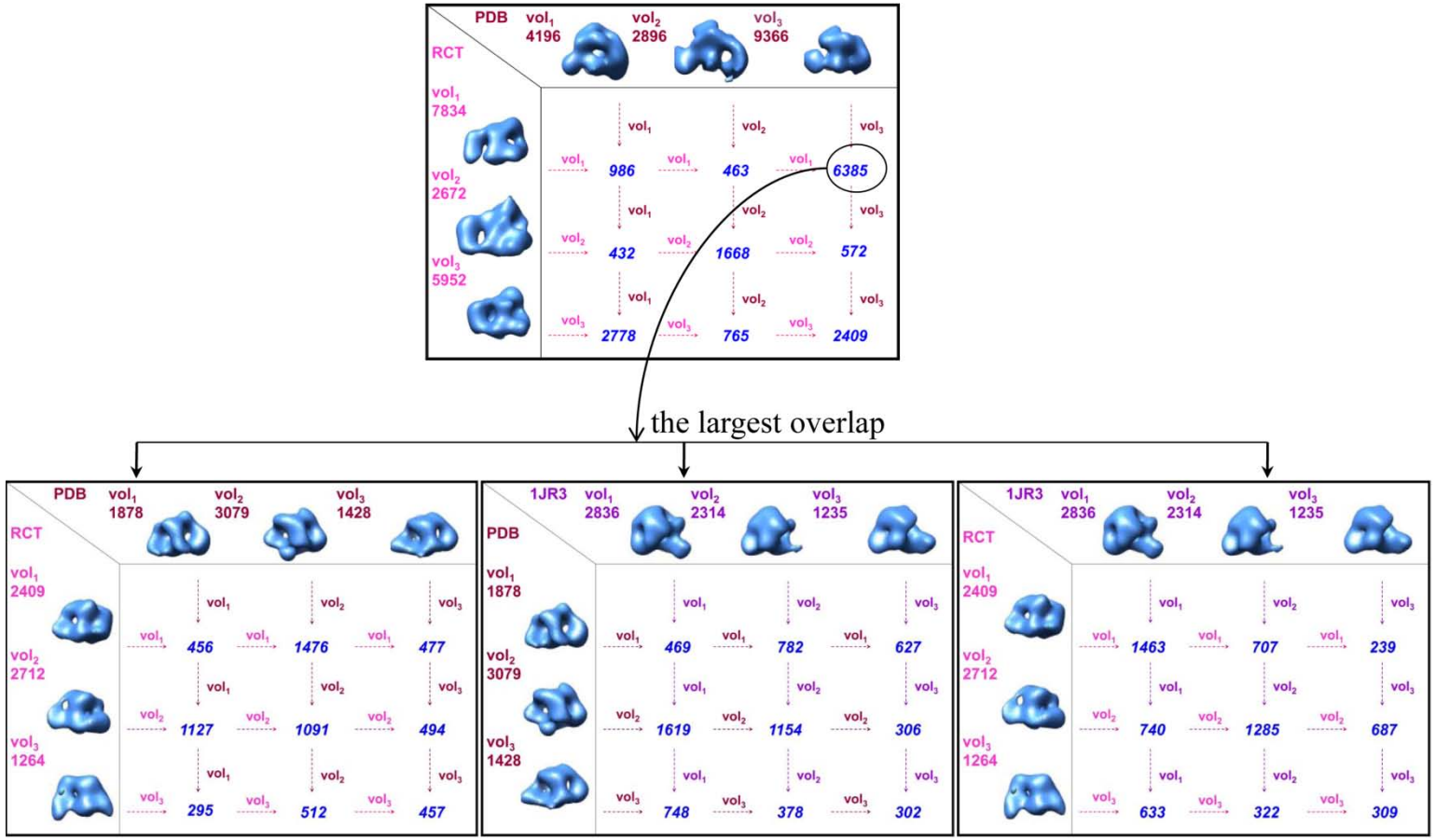


Figure 44. First (upper panel) and second (lower panels) round of MLF3D based dataset cleaning for $\gamma(430)_4\,\psi\chi$. *Violet*, *pink* and *dark red* figures indicate the number of images classified as most likely being projections of each particular seed. Figures *in dark blue* indicate the number of images classified together for each pair of seeds coming from independent MLF3D runs with different initial references. The maps shown are the final results of each MLF3D run.

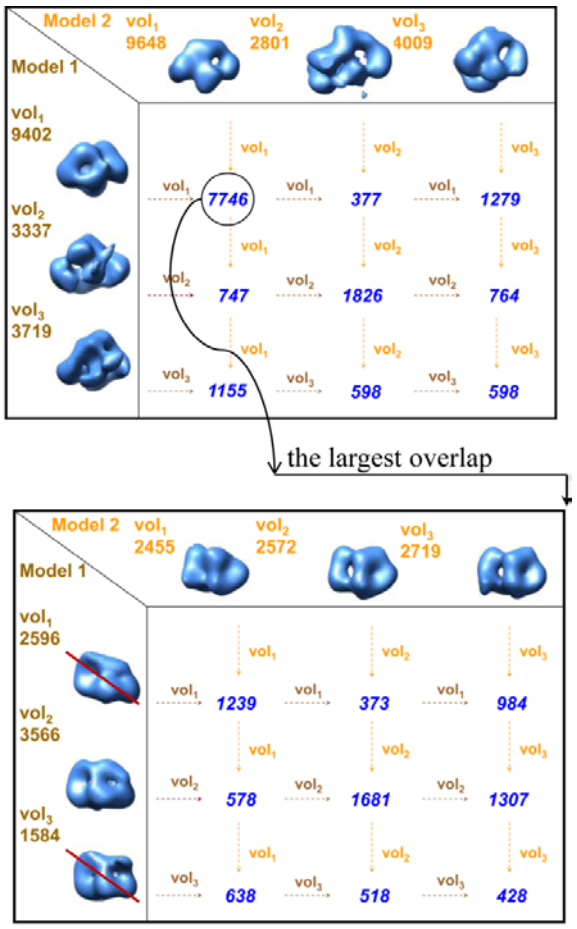


Figure 45. 3rd round of dataset cleaning for γ(430)4ψχ. The complete dataset was used as input to repeat the previous 1st and 2nd rounds with initial maps calculated form a subset of the data Images that went to the crossed out 3D map were not taken to further analysis.

8. *Resumen en español*

8.1 Introducción

El holoenzima DNA polimerasa III (Pol III) de *E. coli* está formado por la polimerasa replicativa (PolIII core), el factor de procesividad β (*clamp* o pinza), y un complejo con actividad ATPasa (complejo DnaX). DnaX realiza múltiples funciones, entre ellas cargar el factor β sobre el DNA (*clamp loading* o cargado de la pinza), y conectar la helicasa a las dos polimerasas catalíticas (una en la hebra líder y otra en la retardada), manteniendo así la coordinación entre las dos hebras durante la replicación del cromosoma. El complejo DnaX está compuesto a su vez por las subunidades δ, δ', ψ, χ, y tres copias del producto del gen *dnaX*, bien en su forma íntegra (proteína τ, 71.1 kDa), o en una forma corta producida por un desplazamiento del marco de lectura del ribosoma (proteína γ, 47.5 kDa). Mientras que un complejo $\gamma_3\delta\delta'$ es suficiente para cargar el factor β sobre el DNA, las subunidades ψ, χ, y particularmente τ son absolutamente necesarias para la supervivencia de la célula, probablemente debido a su papel en la organización del replisoma (Johnson and O'Donnell, 2005).

Los productos del gen *dnaX* pueden formar al menos tres formas oligoméricas estables, dependiendo de la presencia de sus proteínas accesorias. Tanto γ como τ, por sí solos forman homotetrámeros. En presencia de δ y δ', el tetrámero se transforma en un trímero ($\gamma_3\delta\delta'$ ó $\tau_3\delta\delta'$). Mientras que γ_4 no interacciona con el core de PolIII, la misma interacción convierte el tetrámero de τ en un dímero (PolIII-τ-τ-PolIII). Finalmente, el dímero ψχ forma complejos tanto con γ_4 como con τ_4. Todas estas transiciones forman parte del proceso de ensamblaje del holoenzima (McHenry, 2003).

Mediante cristalografía de proteínas se ha resuelto la estructura del *clamp loader* (γ3δδ') mínimo (Jeruzalmi, et al., 2001; Simonetta, et al., 2009), así como del heterodímero ψχ (Kazmirski, et al., 2004) y por el último del complejo $\gamma_3\delta\delta'$ unido a un péptido amino-terminal de ψ (Simonetta, et al., 2009). En este trabajo se analiza la

estructura cuaternaria de los complejos γ_4 y $\gamma_4\psi\chi$ utilizando microscopía electrónica tridimensional.

La microscopia electrónica presenta una serie de ventajas sobre otras técnicas de Biología Estructural: permite trabajar con especímenes de muy diferentes tamaños, requiere de cantidades pequeñas de muestra y además no es necesario que esté muy concentrada. Permite obtener una estructura global de los complejos macromoleculares. Para el estudio de muestras biológicas por microscopía electrónica se utilizan principalmente dos técnicas: la tinción negativa o la crio-microscopía. Las imágenes obtenidas en microscopía electrónica de transmisión son proyecciones bidimensionales de los objetos tridimensionales observados. La información de la tercera dimensión se puede recuperar a partir de la información contenida en cada proyección del objeto, siempre y cuando el conjunto de proyecciones se haya obtenido desde un amplio rango de puntos de vista del objeto tridimensional (Frank 2006).

8.2 Objetivos

El objetivo general de esta tesis es obtener nuevos datos sobre la estructura de los complejos de la polimerasa III lo que podría ayudar a entender el ensamblado dinámico de la enzima. En particular, se han realizado los siguientes objetivos:

- Obtener la mapas 3D de los complejos γ_4 y $\gamma_4\psi\chi$.

- Determinar los diferencias entre contactos proteína-proteína entre subunidades γ en complejos γ_4 (nuestro análisis EM) y $\gamma_3\delta\delta$`(estructure cristalina de mínimum clamp loader) para explicar cómo se produce los cambios $\gamma$ entre tetrámero y trímero en función de la presencia de otras subunidades ($\delta$ and $\delta$`).

- Determinar la organización de región C-terminal, ausente en la estructura cristalina, mediante la asignación de los mapas de diferencia de los homotetramers con la longitud entera y truncada.

- Determinar la localización del heterodimero de $\psi\chi$ mediante los mapas de la diferencia de los complejos cuales la poseen y los que no.

8.3 Materiales y Métodos

La expresión y purificación de los complejos estudiados se realizó en el laboratorio del Dr. Charles S. McHenry (Department of Chemistry and Biochemistry, University of Colorado at Boulder, Colorado, USA). Para caracterizar la composición y estado de oligomerización de los complejos se realizaron ensayos de electroforesis en condiciones desnaturalizantes y cromatografía de exclusión molecular. Para mejorar la preservación y la homogeneidad estructural de los complejos se llevó a cabo una fijación suave con gluteraldehído, bien seguida de cromatografía o combinada con sedimentación en un gradiente de glicerol (GraFix) (Kastner, et al., 2008). Para el análisis estructural de los complejos se ha usado la microscopía electrónica de transmisión y herramientas de procesamiento de imagen (Scheres, et al., 2008). Las muestras se tiñeron con acetato de uranilo al 2% y se tomaron conjuntos de imágenes en condiciones de dosis mínima. Se obtuvieron modelos iniciales a partir de experimentos de random conical tilt (RCT) o estructural cristalinas resueltas previamente. Tanto para el alineamiento y clasificación de las proyecciones en 2D como para la búsqueda de orientaciones y reconstrucción tridimensional se han utilizado algoritmos de máxima verosimilitud en espacio de Fourier (MLF2D ó MLF3D) (Scheres, Gao, Valle, Herman, Eggermont, Frank and Carazo, 2007; Scheres, et al., 2007). El método de máxima verosimilitud permite separar en clases las proyecciones de distintas estructuras (conformaciones) 3D que pueda adoptar la muestra de estudio, a partir de un mapa muy filtrado usado como referencia. También se realizó el análisis de uno de los complejos mediante crio-microscopía electrónica.

8.4 Resultados y Discusión

Aunque los ensayos de electroforesis y exclusión molecular mostraron una composición y estado oligomérico homogéneo de los cuatro complejos estudiados, los análisis por microscopía electrónica demostraban heterogeneidad. La fijación con gluteraldehído seguida de exclusión molecular mejoró la homogeneidad estructural, por lo que se adquirieron conjuntos de imágenes para calcular modelos iniciales por el método de RCT. Sin embargo, cuando tanto estos modelos como otros creados a partir de estructuras cristalográficas se utilizaron para llevar a cabo refinamientos 3D de un nuevo, el proceso no convergía a una solución única. Para intentar mejorar la preservación y homogeneidad estructural de los complejos se utilizó el método GraFix; nuevamente los conjuntos de datos así obtenidos producían diferente mapas 3D dependiendo del modelo inicial utilizado para el refinamiento.

A continuación se intentó seleccionar un subconjunto de imágenes auto-consistente para cada complejo. Se consideró que un subconjunto de imágenes era auto-consistente si las mismas imágenes se agrupaban juntas tras someter el conjunto completo a clasificación por el método MLF3D a partir de diferentes referencias iniciales. De este modo se consiguió obtener mapas 3D consenso para los cuatro complejos estudiados. La validación de los mapas se realizó mediante: a) comparación de sus proyecciones con las imágenes medias obtenidas a partir de clasificaciones 2D (libres de referencia) de las imágenes experimentales; b) comparación con modelos 3D obtenidos por el método de RCT; c) comparación con imágenes medias obtenidas a partir de datos de crio-microscopía electrónica. Este último ensayo indicó que la heterogeneidad estructural observada en muestras preparadas por tinción negativa también estaba presente en muestras vitrificadas, y por tanto no era producida por artefactos del proceso de tinción negativa.

Para interpretar los mapas obtenidos se ajustaron sobre ellos las coordenadas atómicas de cuatro monómeros de γ y del heterodímero ψχ, y se calcularon mapas diferencia entre las reconstrucciones del homotetrámero γ_4 y el complejo $\gamma_4\psi\chi$. La orientación de los monómeros γ dese asignó mediante la comparación con la estructura cristalina del *clamp loader*. Este análisis mostró que en el homotetrámero, γ puede establecer interacciones proteína-proteína con otras copias de sí misma de

múltiples modos, formando una estructura circular asimétrica. Esta versatilidad de interacciones estaría en consonancia con las múltiples actividades e interacciones de γ con otros componentes del holoenzima Pol III. El heterodímero ψχ se inserta en una posición definida dentro del tetrámero, desplazando una de las copias de γ de forma que se reduce su interacción con el resto de monómeros. Esto podría explicar la moderada actividad de ψχ como desestabilizador del tetrámero de γ (Park, et al., 2010).

8.5 Conclusiones

1. La estructura de los compliejos de clamp loade de E.coli γ_4 y$\gamma_4\psi\chi$ se ha analizado mediante microscopía electrónica y procesamiento de imágenes.

2. Una combinación de estabilización química (cross-linking) y la selección estricta de los imagenes (limpieza de los datos) ha sido fundamental para llegar a un consenso sobre las estructuras de estos ejemplares tan difíciles, que son pequeños, asimétricos, y presentar una variabilidad estructural muy alta.

3. Mapas de diferencias permitieron la localización del dímero χ ψ en el γ truncada $(370)_4\psi\chi$, pero no en el complejo de longitud entera γ. Del mismo modo, mapas de diferencias no identificaron las regiones γ C-terminal.

4. La combinación de estructuras cristalinas y mapas EM resultó en un modelo estructural para la $\gamma_4\psi\chi$ γ_4 y complejos.

5. El modelo india que γ tiene una gran versatilidad para establecer las interacciones proteína-proteína, incluso entre los monómeros idénticos. Esta variabilidad es probable que sea relevante para su actividad funcional en la dinámica replisoma.

6. El modelo también muestra cómo ψ χ interactúa con el tetrámero γ en una posición definida, alterando ciertas interacciones entre monómeros γ. Esto podría explicar la limiada actividad para romper el oligómero mostrado previamente en ψ

9. *References*

Anderson, S.G., Williams, C.R., O'Donnell, M. and Bloom, L.B. (2007) A function for the psi subunit in loading the Escherichia coli DNA polymerase sliding clamp, *J Biol Chem*, **282**, 7035-7045.

Baker, T.A. and Bell, S.P. (1998) Polymerases and the replisome: machines within machines, *Cell*, **92**, 295-305.

Blinkova, A., Hervas, C., Stukenberg, P.T., Onrust, R., O'Donnell, M.E. and Walker, J.R. (1993) The Escherichia coli DNA polymerase III holoenzyme contains both products of the dnaX gene, tau and gamma, but only tau is essential, *J Bacteriol*, **175**, 6018-6027.

Blinkowa, A.L. and Walker, J.R. (1990) Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III gamma subunit from within the tau subunit reading frame, *Nucleic Acids Res*, **18**, 1725-1729.

Buchel, C., Morris, E., Orlova, E. and Barber, J. (2001) Localisation of the PsbH subunit in photosystem II: a new approach using labelling of His-tags with a Ni(2+)-NTA gold cluster and single particle analysis, *J Mol Biol*, **312**, 371-379.

Crowther, R.A., DeRosier, D. J. & Klug, A. (1970) The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy., *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences.* , **317**, 319-340.

Chen, J.Z., Settembre, E.C., Aoki, S.T., Zhang, X., Bellamy, A.R., Dormitzer, P.R., Harrison, S.C. and Grigorieff, N. (2009) Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM, *Proc Natl Acad Sci U S A*, **106**, 10644-10648.

Chiu, W. (1993) What does electron cryomicroscopy provide that X-ray crystallography and NMR spectroscopy cannot?, *Annu Rev Biophys Biomol Struct*, **22**, 233-255.

Dallmann, H.G., Kim, S., Pritchard, A.E., Marians, K.J. and McHenry, C.S. (2000) Characterization of the unique C terminus of the Escherichia coli tau DnaX protein. Monomeric C-tau binds alpha AND DnaB and can partially replace tau in reconstituted replication forks, *J Biol Chem*, **275**, 15512-15519.

Dallmann, H.G. and McHenry, C.S. (1995) DnaX complex of Escherichia coli DNA polymerase III holoenzyme. Physical characterization of the DnaX subunits and complexes, *J Biol Chem*, **270**, 29563-29569.

Dallmann, H.G., Thimmig, R.L. and McHenry, C.S. (1995) DnaX complex of Escherichia coli DNA polymerase III holoenzyme. Central role of tau in initiation complex assembly and in determining the functional asymmetry of holoenzyme, *J Biol Chem*, **270**, 29555-29562.

Davey, M.J., Jeruzalmi, D., Kuriyan, J. and O'Donnell, M. (2002) Motors and switches: AAA+ machines within the replisome, *Nat Rev Mol Cell Biol*, **3**, 826-835.

Duggin, I.G., Wake, R.G., Bell, S.D. and Hill, T.M. (2008) The replication fork trap and termination of chromosome replication, *Mol Microbiol*, **70**, 1323-1333.

Flower, A.M. and McHenry, C.S. (1990) The gamma subunit of DNA polymerase III holoenzyme of Escherichia coli is produced by ribosomal frameshifting, *Proc Natl Acad Sci U S A*, **87**, 3713-3717.

Frank, J. (2006) Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State *Oxford University Press, New York*, .

Gao, D. and McHenry, C.S. (2001a) tau binds and organizes Escherichia coli replication proteins through distinct domains. Domain IV, located within the unique C terminus of tau, binds the replication fork, helicase, DnaB, *J Biol Chem*, **276**, 4441-4446.

Gao, D. and McHenry, C.S. (2001b) tau binds and organizes Escherichia coli replication through distinct domains. Partial proteolysis of terminally tagged tau to determine candidate domains and to assign domain V as the alpha binding domain, *J Biol Chem*, **276**, 4433-4440.

Gao, D. and McHenry, C.S. (2001c) Tau binds and organizes Escherichia coli replication proteins through distinct domains. Domain III, shared by gamma and tau, binds delta delta ' and chi psi, *J Biol Chem*, **276**, 4447-4453.

Glaeser, R.M. (2008) Macromolecular structures without crystals, *Proc Natl Acad Sci U S A*, **105**, 1779-1780.

Glover, B.P. and McHenry, C.S. (1998) The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template, *J Biol Chem*, **273**, 23476-23484.

Goedken, E.R., Levitus, M., Johnson, A., Bustamante, C., O'Donnell, M. and Kuriyan, J. (2004) Fluorescence measurements on the E.coli DNA polymerase clamp loader: implications for conformational changes during ATP and clamp binding, *J Mol Biol*, **336**, 1047-1059.

Gómez-Llorente, Y. (2007) Estudios estructurales en la proteína helicasa MCM de *Methanobacterium thermoautotrophicum.*, *Tesis doctoral*.

Grigorieff, N. (2000) Resolution measurement in structures derived from single particles, *Acta Crystallogr D Biol Crystallogr*, **56**, 1270-1277.

Guenther, B., Onrust, R., Sali, A., O'Donnell, M. and Kuriyan, J. (1997) Crystal structure of the delta' subunit of the clamp-loader complex of E. coli DNA polymerase III, *Cell*, **91**, 335-345.

Gulbis, J.M., Kazmirski, S.L., Finkelstein, J., Kelman, Z., O'Donnell, M. and Kuriyan, J. (2004) Crystal structure of the chi:psi sub-assembly of the Escherichia coli DNA polymerase clamp-loader complex, *Eur J Biochem*, **271**, 439-449.

Henderson, R. (2004) Realizing the potential of electron cryo-microscopy, *Q Rev Biophys*, **37**, 3-13.

Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kuhne, R., Stout, J.R., Higman, V.A., Klevit, R.E., van Rossum, B.J. and Oschkinat, H. (2010) Solid-state NMR and SAXS studies provide a structural basis for the activation of alphaB-crystallin oligomers, *Nat Struct Mol Biol*, **17**, 1037-1042.

Jeruzalmi, D., O'Donnell, M. and Kuriyan, J. (2001) Crystal structure of the processivity clamp loader gamma (gamma) complex of E. coli DNA polymerase III, *Cell*, **106**, 429-441.

Johnson, A. and O'Donnell, M. (2005) Cellular DNA replicases: components and dynamics at the replication fork, *Annu Rev Biochem*, **74**, 283-315.

Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., Uchtenhagen, H., Urlaub, H., Herzog, F., Peters, J.M., Poerschke, D., Luhrmann, R. and Stark, H. (2008) GraFix: sample preparation for single-particle electron cryomicroscopy, *Nat Methods*, **5**, 53-55.

Kazmirski, S.L., Podobnik, M., Weitze, T.F., O'Donnell, M. and Kuriyan, J. (2004) Structural analysis of the inactive state of the Escherichia coli DNA polymerase clamp-loader complex, *Proc Natl Acad Sci U S A*, **101**, 16750-16755.

Kim, S., Dallmann, H.G., McHenry, C.S. and Marians, K.J. (1996) Coupling of a replicative polymerase and helicase: a tau-DnaB interaction mediates rapid replication fork movement, *Cell*, **84**, 643-650.

Kornberg, A., Baker, T. (1992) DNA Replication.

Leschziner, A.E. and Nogales, E. (2007) Visualizing flexibility at molecular resolution: analysis of heterogeneity in single-particle electron microscopy reconstructions, *Annu Rev Biophys Biomol Struct*, **36**, 43-62.

Leu, F.P. and O'Donnell, M. (2001) Interplay of clamp loader subunits in opening the beta sliding clamp of Escherichia coli DNA polymerase III holoenzyme, *J Biol Chem*, **276**, 47185-47194.

Liu, H., Jin, L., Koh, S.B., Atanasov, I., Schein, S., Wu, L. and Zhou, Z.H. (2010) Atomic structure of human adenovirus by cryo-EM reveals interactions among protein networks, *Science*, **329**, 1038-1043.

Lovett, S.T. (2007) Polymerase switching in DNA replication, *Mol Cell*, **27**, 523-526.

Ludtke, S.J., Baldwin, P.R. and Chiu, W. (1999) EMAN: semiautomated software for high-resolution single-particle reconstructions, *J Struct Biol*, **128**, 82-97.

Martinez-Jimenez, M.I., Mesa, P. and Alonso, J.C. (2002) Bacillus subtilis tau subunit of DNA polymerase III interacts with bacteriophage SPP1 replicative DNA helicase G40P, *Nucleic Acids Res*, **30**, 5056-5064.

McInerney, P., Johnson, A., Katz, F. and O'Donnell, M. (2007) Characterization of a triple DNA polymerase replisome, *Mol Cell*, **27**, 527-538.

McHenry, C.S. (2003) Chromosomal replicases as asymmetric dimers: studies of subunit arrangement and functional consequences, *Mol Microbiol*, **49**, 1157-1165.

Mindell, J.A. and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy, *J Struct Biol*, **142**, 334-347.

Miyata, T., Oyama, T., Mayanagi, K., Ishino, S., Ishino, Y. and Morikawa, K. (2004) The clamp-loading complex for processive DNA replication, *Nat Struct Mol Biol*, **11**, 632-636.

Nuñez-Ramirez, R., Velten, M., Rivas, G., Polard, P., Carazo, J.M. and Donate, L.E. (2007) Loading a ring: structure of the Bacillus subtilis DnaB protein, a co-loader of the replicative helicase, *J Mol Biol*, **367**, 764-769.

O'Donnell, M., Jeruzalmi, D. and Kuriyan, J. (2001) Clamp loader structure predicts the architecture of DNA polymerase III holoenzyme and RFC, *Curr Biol*, **11**, R935-946.

Olson, M.W., Dallmann, H.G. and McHenry, C.S. (1995) DnaX complex of Escherichia coli DNA polymerase III holoenzyme. The chi psi complex functions by increasing the affinity of tau and gamma for delta.delta' to a physiologically relevant range, *J Biol Chem*, **270**, 29570-29577.

Park, A.Y., Jergic, S., Politis, A., Ruotolo, B.T., Hirshberg, D., Jessop, L.L., Beck, J.L., Barsky, D., O'Donnell, M., Dixon, N.E. and Robinson, C.V. (2010) A single subunit directs the assembly of the Escherichia coli DNA sliding clamp loader, *Structure*, **18**, 285-292.

Penczek, P.A. (2002) Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms, *J Struct Biol*, **138**, 34-46.

Penczek, P.A., Grassucci, R.A. and Frank, J. (1994) The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles, *Ultramicroscopy*, **53**, 251-270.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem*, **25**, 1605-1612.

Pieper, U., Eswar, N., Webb, B.M., Eramian, D., Kelly, L., Barkan, D.T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M.A., Davis, F.P. and Sali, A. (2009) MODBASE, a database of annotated comparative protein structure models and associated resources, *Nucleic Acids Res*, **37**, D347-354.

Podobnik, M., Weitze, T.F., O'Donnell, M. and Kuriyan, J. (2003) Nucleotide-induced conformational changes in an isolated Escherichia coli DNA polymerase III clamp loader subunit, *Structure*, **11**, 253-263.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I. and Sussman, J.L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded, *Bioinformatics*, **21**, 3435-3438.

Pritchard, A.E., Dallmann, H.G., Glover, B.P. and McHenry, C.S. (2000) A novel assembly mechanism for the DNA polymerase III holoenzyme DnaX complex: association of deltadelta' with DnaX(4) forms DnaX(3)deltadelta', *Embo J*, **19**, 6536-6545.

Pritchard, A.E. and McHenry, C.S. (2001) Assembly of DNA polymerase III holoenzyme: co-assembly of gamma and tau is inhibited by DnaX complex accessory proteins but stimulated by DNA polymerase III core, *J Biol Chem*, **276**, 35217-35222.

Putnam, C.D., Hammel, M., Hura, G.L. and Tainer, J.A. (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, *Q Rev Biophys*, **40**, 191-285.

Radermacher, M. (1988) Three-dimensional reconstruction of single particles from random and nonrandom tilt series, *J Electron Microsc Tech*, **9**, 359-394.

Reyes-Lamothe, R., Sherratt, D.J. and Leake, M.C. (2010) Stoichiometry and architecture of active DNA replication machinery in Escherichia coli, *Science*, **328**, 498-501.

Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J. and Carazo, J.M. (2007) Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization, *Nat Methods*, **4**, 27-29.

Scheres, S.H., Nuñez-Ramirez, R., Gomez-Llorente, Y., San Martin, C., Eggermont, P.P. and Carazo, J.M. (2007) Modeling experimental image formation for likelihood-based classification of electron microscopy data, *Structure*, **15**, 1167-1177.

Scheres, S.H., Nuñez-Ramirez, R., Sorzano, C.O., Carazo, J.M. and Marabini, R. (2008) Image processing for electron microscopy single-particle analysis using XMIPP, *Nat Protoc*, **3**, 977-990.

Shaikh, T.R., Hegerl, R. and Frank, J. (2003) An approach to examining model dependence in EM reconstructions using cross-validation, *J Struct Biol*, **142**, 301-310.

Shigematsu, H., Sokabe, T., Danev, R., Tominaga, M. and Nagayama, K. (2010) A 3.5-nm structure of rat TRPV4 cation channel revealed by Zernike phase-contrast cryoelectron microscopy, *J Biol Chem*, **285**, 11210-11218.

Siebert, X. and Navaza, J. (2009) UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions, *Acta Crystallogr D Biol Crystallogr*, **65**, 651-658.

Simonetta, K.R., Kazmirski, S.L., Goedken, E.R., Cantor, A.J., Kelch, B.A., McNally, R., Seyedin, S.N., Makino, D.L., O'Donnell, M. and Kuriyan, J. (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex, *Cell*, **137**, 659-671.

Spahn, C.M. and Penczek, P.A. (2009) Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM, *Curr Opin Struct Biol*, **19**, 623-631.

Subramaniam, S. and Henderson, R. (2000) Molecular mechanism of vectorial proton translocation by bacteriorhodopsin, *Nature*, **406**, 653-657.

Tidow, H., Melero, R., Mylonas, E., Freund, S.M., Grossmann, J.G., Carazo, J.M., Svergun, D.I., Valle, M. and Fersht, A.R. (2007) Quaternary structures of tumor suppressor p53 and a specific p53 DNA complex, *Proc Natl Acad Sci U S A*, **104**, 12324-12329.

Tsuchihashi, Z. and Kornberg, A. (1990) Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme, *Proc Natl Acad Sci U S A*, **87**, 2516-2520.

Walther, P. (2008) High-resolution cryo-SEM allows direct identification of F-actin at the inner nuclear membrane of Xenopus oocytes by virtue of its structural features, *J Microsc*, **232**, 379-385.

Xiao, H., Dong, Z. and O'Donnell, M. (1993) DNA polymerase III accessory proteins. IV. Characterization of chi and psi, *J Biol Chem*, **268**, 11779-11784.

Yonekura, K., Maki-Yonekura, S. and Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy, *Nature*, **424**, 643-650.

Yu, X., Jin, L. and Zhou, Z.H. (2008) 3.88 A structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy, *Nature*, **453**, 415-419.

10. *Glossary*

Pol III	*DNA polymerase III holoenzyme*
ATPγS	*Adenosine 5'-(gamma-thiotriphosphate. One of the gamma-phosphate oxygens is replaced by a sulfur atom.*
AAA+	*ATPases Associated with diverse cellular Activities. They share a common conserved module of approximately 230 amino acid residues.*
SSB	*Single-stranded DNA binding proteins*
ATP	*Adenosine triphosphate*
DNA	*Deoxyribonucleic acid*
DTT	*Dithiothreitol [*
EDTA	*Ethylenediamine tetraacetic acid*
RNA	*Ribonucleic acid*
TRIS	*Tris(hydroxymethyl)aminomethane*
NMR	*Nuclear magnetic resonance*
GL	glutaraldehyde
SAXS	small angle X-ray scattering
3DEM	*Three Dimensional Electron Microscopy*
Cryo-EM	*Cryo-electron microscopy*
X-Ray	*X-radiation*
2D	*Two Dimensional*
3D	*Three dimensional*
RCT	*Random Conical Tilt*
EM	*Electron Microscopy*
PDB	*Protein Data Bank* www.pdb.org
OD	*Optical Density*
MLF2D	*Two dimensional-classification by maximum-likelihood in Fourier space*
MLF3D	*Three dimensional classification by maximum-likelihood in Fourier space*
SSNR	*Spectral Signal to Noise Ratio*
Ni-NTA	*Nickel-nitrilotriacetic acid*
E. coli	*Escherichia coli*
B.subtilis	*Bacillus subtilis*
EMAN	http://blake.bcm.tmc.edu/eman/
XMIPP	*X-Window-based Microscopy Image Processing Package* http://xmipp.cnb.csic.es/
UROX	http://mem.ibs.fr/UROX
UCSF CHIMERA	www.cgl.ucsf.edu/chimera/
Fei Company	www.fei.com
Gatan, Inc	www.gatan.com
Jeol Ltd	www.jeol.com
Zeiss; Carl Zeiss, Inc.	www.zeiss.com
BioRad	www.bio-rad.com
Sigma Aldrich	www.**sigmaaldrich**.com
MW	*Molecular Weight*
px	*Pixel*
Mpx	*Mega pixel*
kDa	*kilo Dalton*
nm	*nano meter*
kV	*kilo Volts*
µm	*micro meter*
Å	*Angstrom*